

1000

```

Query Match      22.1%; Score 60; DB 15; Length 70;
Best Local Similarity 52.4%; Pred. No. 2.3;
Matches 11; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 31 WPSQEQGNFSLSRKLRDG 51
DB 39 YPGQHGTEPSLKIKQVRCG 59

RESULT 10
US-09-468-147-91
; Sequence 91, Application US/09468147A
; Publication No. US20030049601A1
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Schlauder, George G.
; APPLICANT: Erker, James C.
; APPLICANT: Desai, Suresh M.
; APPLICANT: Dawson, George J.
; APPLICANT: Washawar, I. K.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTING
; TITLE OF INVENTION: HEPATITIS E VIRUS
; FILE REFERENCE: 6232 US.P1
; CURRENT APPLICATION NUMBER: US/09/468,147A
; CURRENT FILING DATE: 1999-12-21
; EARLIER APPLICATION NUMBER: US 09/173,141
; EARLIER FILING DATE: 1998-10-15
; EARLIER APPLICATION NUMBER: US 60/061,199
; EARLIER FILING DATE: 1997-10-15
; NUMBER OF SEQ ID NOS: 258
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 91
; LENGTH: 1698
; TYPE: PRT
; ORGANISM: Hepatitis E Virus
; FEATURE:
; OTHER INFORMATION: Xaa = Unknown or Other at position 174
; OTHER INFORMATION: Xaa = Unknown or Other at position 363
; OTHER INFORMATION: Xaa = Unknown or Other at position 1088
; OTHER INFORMATION: Xaa = Unknown or Other at position 1131
; OTHER INFORMATION: Xaa = Unknown or Other at position 1217
; OTHER INFORMATION: Xaa = Unknown or Other at position 1389
US-09-468-147-91

Query Match      22.0%; Score 59.5; DB 11; Length 1698;
Best Local Similarity 44.4%; Pred. No. 1.1e+02;
Matches 12; Conservative 4; Mismatches 10; Indels 1; Gaps 1;

QY 3 HNEGFKLFATEATSDWLNANNVPATPV 29
DB 911 HRPQDELYLTPAANWFANK-PAQPV 936

RESULT 11
US-09-285-385C-4
; Sequence 4, Application US/09285385C
; Patent No. US20020150946A1
; GENERAL INFORMATION:
; APPLICANT: Greenspan, Daniel S.
; APPLICANT: Scott, Ian C.
; APPLICANT: Thomas, Christina L.
; TITLE OF INVENTION: MAMMALIAN TOLLOID-LIKE GENE AND PROTEIN
; FILE REFERENCE: 960296.96111
; CURRENT APPLICATION NUMBER: US/09/285,385C
; CURRENT FILING DATE: 1999-04-02
; PRIOR APPLICATION NUMBER: 60/111973
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 60/080550
; PRIOR FILING DATE: 1998-04-03
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 4
; LENGTH: 1018

Query Match      21.8%; Score 59; DB 10; Length 1012;
Best Local Similarity 25.4%; Pred. No. 69;
Matches 15; Conservative 7; Mismatches 25; Indels 12; Gaps 2;

QY 2 LHNEGFKLFATEATSDWL-----NANNVPATPVAVPSQEQGNFSLSRKLRDG 51
DB 65 LDEDDLKLFHIDRAEDWTKPSIDKPGHDTGGLETSARWPNFASNAIQAPRK---DG 120

RESULT 12
US-10-029-180-102
; Sequence 102, Application US/10029180
; Publication No. US20020182708A1
; GENERAL INFORMATION:
; APPLICANT: Cali, Brian M.
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin T.
; APPLICANT: Milna, G. Todd
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeffrey C.
; APPLICANT: Trueheart, Josh
; APPLICANT: Zhang, Lixin
; TITLE OF INVENTION: No. US20020182708A1el Regulators of Fungal Gene Expression
; FILE REFERENCE: MIC-004
; CURRENT APPLICATION NUMBER: US/10/029,180
; CURRENT FILING DATE: 2001-12-22
; PRIOR APPLICATION NUMBER: US 60/257,431
; PRIOR FILING DATE: 2000-12-22
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 102
; LENGTH: 576
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fungal gene
US-10-029-180-102

Query Match      21.2%; Score 57.5; DB 14; Length 576;
Best Local Similarity 44.8%; Pred. No. 56;
Matches 13; Conservative 3; Mismatches 12; Indels 1; Gaps 1;

QY 15 TSDWLNANNVPATPVAVPSQEQGNFSLS 42
DB 96 TPDWSLQNTVPMPVATFPSSSTSPGLT 124

RESULT 13
US-09-494-359-3
; Sequence 3, Application US/09494359
; Publication No. US20030124685A1
; GENERAL INFORMATION:
; APPLICANT: Kuwabara, Yoko
; APPLICANT: HASHIGUCHI, Kenichi
; APPLICANT: NAKAMATSU, Tsuyoshi
; APPLICANT: KURAHASHI, Osamu
; APPLICANT: MORI, Yukiko
; APPLICANT: ITO, Hisao
; TITLE OF INVENTION: CARBAMOYL-PHOSPHATE SYNTHETASE GENE OF CORNEIFORM BACTERIA
; TITLE OF INVENTION: PRODUCING L-ARGININE
; FILE REFERENCE: 0010-1082-0
; CURRENT APPLICATION NUMBER: US/09/494,359
; CURRENT FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: JP 11-24149
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 1018

```

;
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,964
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/241,809
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/249,299
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/236,327
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/241,785
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/244,617
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 60/225,268
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/236,368
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/251,856
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/251,868
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/229,344
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,287
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,513
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: 60/231,413
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/229,509
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: 60/236,367
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/237,039
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,038
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/236,370
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/236,802
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,037
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,040
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/240,960
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/239,935
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/239,937
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/241,787
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,474
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/246,532
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/249,216
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,210
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/226,681
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/225,759
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/225,213
; PRIOR FILING DATE: 2000-08-14
;
; PRIOR APPLICATION NUMBER: 60/227,182
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/225,214
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/235,836
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/230,438
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/215,135
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 60/225,266
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/249,218
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,208
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,213
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,212
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,207
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,245
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,244
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,217
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,211
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,215
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,264
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,214
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,297
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/232,400
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/231,242
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,081
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,080
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,414
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,244
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/233,064
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/233,063
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,397
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,399
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,401
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/241,808
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,826
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,786
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,221
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,475
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/231,243
; PRIOR FILING DATE: 2000-09-08

```
QY 2 LHNQGFPLFATEATSWLNANNVPATPV--AWFSQEQNPSSLSSIR 45
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 28 VYNHAFRL-----TADWSTAEYDMMATTFLEAWRLRERVDPGGSLR 68

RESULT 6
US-09-764-878-129
; Sequence 129, Application US/09764878
; Patent No. US20020090615A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PAL21
; CURRENT APPLICATION NUMBER: US/09/764,878
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 428
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 129
; LENGTH: 70
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-878-129

Query Match 22.1%; Score 60; DB 9; Length 70;
Best Local Similarity 52.4%; Pred. No. 2.3;
Matches 11; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 31 WPSQEQGNPSSLSSIRKLIRDG 51
      :|||::|||::|||::|||::|||
Db 39 YPGQHGTPSLKIQKLVRG 59

RESULT 7
US-09-764-860-464
; Sequence 464, Application US/09764860
; Patent No. US20020094953A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC008
; CURRENT APPLICATION NUMBER: US/09/764,860
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1198
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 464
; LENGTH: 70
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-860-464

Query Match 22.1%; Score 60; DB 9; Length 70;
Best Local Similarity 52.4%; Pred. No. 2.3;
Matches 11; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 31 WPSQEQGNPSSLSSIRKLIRDG 51
      :|||::|||::|||::|||::|||
Db 39 YPGQHGTPSLKIQKLVRG 59

RESULT 8
US-10-079-854-129
; Sequence 129, Application US/10079854
; Publication No. US20030054368A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PAL21C1
; CURRENT APPLICATION NUMBER: US/10/079,854
; CURRENT FILING DATE: 2002-02-22
; Prior application removed - See file Wrapper or Palm
; NUMBER OF SEQ ID NOS: 428
```

```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 129
; LENGTH: 70
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-079-854-129

Query Match 22.1%; Score 60; DB 15; Length 70;
Best Local Similarity 52.4%; Pred. No. 2.3;
Matches 11; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 31 WPSQEQGNPSSLSSIRKLIRDG 51
      :|||::|||::|||::|||::|||
Db 39 YPGQHGTPSLKIQKLVRG 59

RESULT 9
US-10-074-095-464
; Sequence 464, Application US/10074095
; Publication No. US2003007704A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC008C1
; CURRENT APPLICATION NUMBER: US/10/074,095
; CURRENT FILING DATE: 2002-02-14
; PRIOR APPLICATION NUMBER: 09/764,860
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/225,757
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/226,868
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/216,647
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/225,267
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/216,880
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/225,270
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/251,869
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/235,834
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/234,274
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: 60/234,223
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: 60/228,924
; PRIOR FILING DATE: 2000-08-30
; PRIOR APPLICATION NUMBER: 60/224,518
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/236,369
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/224,519
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2003, 02:16:53 ; Search time 36 Seconds
(without alignments)
206.709 Million cell updates/sec

Title: US-09-585-077C-4_COPY_1380_1430
Perfect score: 271
Sequence: 1 QLNHEGFKLFATEATSDWLN.....PSQEGNPSSLIRKLRDG 51

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 541936 seqs, 145912426 residues

Total number of hits satisfying chosen parameters: 541936

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2.6/ptodata/2/pubpaa/US07_PUBCOMB.pap.*
- 2: /cgn2.6/ptodata/2/pubpaa/PCT_NEW_PUB.pap.*
- 3: /cgn2.6/ptodata/2/pubpaa/US06_NEW_PUB.pap.*
- 4: /cgn2.6/ptodata/2/pubpaa/US06_PUBCOMB.pap.*
- 5: /cgn2.6/ptodata/2/pubpaa/US07_NEW_PUB.pap.*
- 6: /cgn2.6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pap.*
- 7: /cgn2.6/ptodata/2/pubpaa/US08_NEW_PUB.pap.*
- 8: /cgn2.6/ptodata/2/pubpaa/US08_PUBCOMB.pap.*
- 9: /cgn2.6/ptodata/2/pubpaa/US09A_PUBCOMB.pap.*
- 10: /cgn2.6/ptodata/2/pubpaa/US09B_PUBCOMB.pap.*
- 11: /cgn2.6/ptodata/2/pubpaa/US09C_PUBCOMB.pap.*
- 12: /cgn2.6/ptodata/2/pubpaa/US09_NEW_PUB.pap.*
- 13: /cgn2.6/ptodata/2/pubpaa/US10A_PUBCOMB.pap.*
- 14: /cgn2.6/ptodata/2/pubpaa/US10B_PUBCOMB.pap.*
- 15: /cgn2.6/ptodata/2/pubpaa/US10C_PUBCOMB.pap.*
- 16: /cgn2.6/ptodata/2/pubpaa/US10_NEW_PUB.pap.*
- 17: /cgn2.6/ptodata/2/pubpaa/US60_NEW_PUB.pap.*
- 18: /cgn2.6/ptodata/2/pubpaa/US60_PUBCOMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	71	26.2	1102	15	US-10-156-761-14395
2	69.5	25.6	344	9	US-09-815-242-5059
3	64	23.6	339	15	US-10-169-048-40
4	62	22.9	619	10	US-09-976-059-30
5	60.5	22.3	232	15	US-10-156-761-11422
6	60	22.1	70	9	US-09-764-878-129
7	60	22.1	70	9	US-09-764-860-464
8	60	22.1	70	15	US-10-079-854-129
9	60	22.1	70	15	US-10-074-095-464
10	59.5	22.0	1698	11	US-09-468-147-91
11	59	21.8	1012	10	US-09-285-385C-4
12	57.5	21.2	576	14	US-10-029-180-102
13	57	21.0	1018	11	US-09-494-359-3
14	57	21.0	1113	9	US-09-836-470B-3
15	57	21.0	1113	15	US-10-284-138-3

16	57	21.0	1113	15	US-10-284-334-3	Sequence 3, Appli
17	56.5	20.8	504	15	US-10-156-761-12417	Sequence 12417, A
18	56.5	20.8	673	15	US-10-157-031-291	Sequence 291, App
19	55.5	20.5	823	15	US-10-177-293-240	Sequence 240, App
20	55.5	20.5	1435	15	US-10-128-714-8125	Sequence 8125, Ap
21	55	20.3	441	9	US-09-879-957-34	Sequence 34, Appl
22	55	20.3	1113	10	US-09-738-626-5279	Sequence 5279, Ap
23	54.5	20.1	1228	10	US-09-117-447-2	Sequence 2, Appli
24	54	19.9	100	10	US-09-872-523-8	Sequence 8, Appli
25	54	19.9	101	15	US-10-097-065-576	Sequence 576, App
26	54	19.9	140	12	US-10-238-075-984	Sequence 984, App
27	54	19.9	570	9	US-09-815-242-10760	Sequence 10760, A
28	54	19.9	3739	10	US-09-861-289-33	Sequence 33, Appl
29	54	19.9	3739	10	US-09-860-846-33	Sequence 33, Appl
30	54	19.9	3739	11	US-09-988-384B-33	Sequence 33, Appl
31	54	19.9	3739	11	US-09-836-821-33	Sequence 33, Appl
32	54	19.9	3739	11	US-09-793-708-2	Sequence 2, Appli
33	54	19.9	3739	12	US-10-201-365-3	Sequence 3, Appli
34	54	19.9	3739	12	US-10-160-539-2	Sequence 2, Appli
35	54	19.9	11877	10	US-09-861-289-6	Sequence 6, Appli
36	54	19.9	11877	10	US-09-860-846-6	Sequence 6, Appli
37	54	19.9	11877	11	US-09-836-821-6	Sequence 6, Appli
38	54	19.9	12199	11	US-09-988-384B-6	Sequence 6, Appli
39	53.5	19.7	311	15	US-10-156-761-14936	Sequence 14936, A
40	53.5	19.7	639	10	US-09-906-209-2	Sequence 2, Appli
41	53	19.6	276	14	US-10-033-078-16	Sequence 16, Appl
42	53	19.6	288	15	US-10-203-823-194	Sequence 194, App
43	53	19.6	401	9	US-09-062-113-63	Sequence 63, Appl
44	53	19.6	401	12	US-10-232-858-63	Sequence 63, Appl
45	53	19.6	452	12	US-10-238-075-859	Sequence 859, App

ALIGNMENTS

RESULT 1
US-10-156-761-14395
; Sequence 14395, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 14395
; LENGTH: 1102
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-14395

Query Match 26.2%; Score 71; DB 15; Length 1102;
Best Local Similarity 40.4%; Pred. No. 2.2;
Matches 21; Conservative 8; Mismatches 21; Indels 2; Gaps 2;

QY 1 QLNHEGFKLFATEATSDWLNANNVPATVPVWPSQEQNP-SLSSIRKLRDG 51

Db 981 ELVAHFELLATSGTAELVKRNGINATVVRKQS-EEGPGGEKTVQLIHGG 1031

RESULT 2

US-09-815-242-5059

; APPLICANT: Sakowicz, Roman
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/09/723,820
; CURRENT FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 09/541,782
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1073
; TYPE: PRT
; ORGANISM: S.pombe
US-09-723-820-6

Query Match 19.9%; Score 54; DB 4; Length 1073;
Best Local Similarity 31.4%; Pred. No. 1.6e+02;
Matches 16; Conservative 8; Mismatches 21; Indels 6; Gaps 2;

QY 5 EGFKLFATEATSDNLNANNVPATPVNPS-----OEGQNPFLSSIRKLIRD 50
| | | | | : | | | | : | | | | : | | | | :
Db 986 EDTSLVKLETTGDTPSKRELPA TP-SWTRDSSLIKETTINLNLDSKKFYRE 1035

Search completed: September 13, 2003, 02:17:36
Job time : 22 secs

APPLICANT: SPARKS, Andrew B.
APPLICANT: HOFFMAN, No. 6309820h
APPLICANT: KAY, Brian K.
APPLICANT: FOWLKES, Dana M.
APPLICANT: MCCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
TITLE OF INVENTION: USING SAME
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,915A
FILING DATE: 03-APR-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Mistock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 441 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-630-915A-34
Query Match 20.3%; Score 55; DB 4; Length 441;
Best Local Similarity 37.5%; Pred. No. 38;
Matches 9; Conservative 3; Mismatches 12; Indels 0; Gaps 0;
QY 18 WLNANNVATVPVPSQEGNPSL 41
Db 283 WFRSTSGCPMPNPPQEEWNPDL 306
RESULT 13
US-09-107-532A-6689
Sequence 6689, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Arianiello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 6689:
SEQUENCE CHARACTERISTICS:
LENGTH: 569 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc-feature
LOCATION: (B) LOCATION 1...569
SEQUENCE DESCRIPTION: SEQ ID NO: 6689:
US-09-107-532A-6689
Query Match 19.9%; Score 54; DB 4; Length 569;
Best Local Similarity 23.9%; Pred. No. 71;
Matches 11; Conservative 12; Mismatches 23; Indels 0; Gaps 0;
QY 2 LHNEGKLPATEATSDWLNANNVATVPVPSQEGNPSLSIRKL 47
Db 466 LQDQLTMKTAKTDTYWLHAKNPGSHVIKSKDKPSDETITEAEL 511
RESULT 14
US-09-541-782-6
Sequence 6, Application US/09541782
Patent No. 6284480
GENERAL INFORMATION:
APPLICANT: Nislow, Corey
APPLICANT: Sakowicz, Roman
APPLICANT: Berand, Christophe
TITLE OF INVENTION: Antifungal Assay
FILE REFERENCE: 1015
CURRENT APPLICATION NUMBER: US/09/541,782
CURRENT FILING DATE: 2000-04-03
NUMBER OF SEQ ID NOS: 10
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6
LENGTH: 1073
TYPE: PRT
ORGANISM: S.pombe
US-09-541-782-6
Query Match 19.9%; Score 54; DB 3; Length 1073;
Best Local Similarity 31.4%; Pred. No. 1.6e+02;
Matches 16; Conservative 8; Mismatches 21; Indels 6; Gaps 2;
QY 5 EGFKLPATEATSDWLNANNVATVPVPSQEGNPSLSIRKL 50
Db 986 EDTSLVLETTGDTPSKRELPA TP-SWTRDSSLLIKETTNLDSPKFEVRE 1035
RESULT 15
US-09-723-820-6
Sequence 6, Application US/09723820
Patent No. 6468760
GENERAL INFORMATION:
APPLICANT: Nislow, Corey


```
Best Local Similarity 98.0%; Pred. No. 2.5e-26;
Matches 50; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 QLNHEGFKLFATEATSDWLNANNVPATPVAMPVPSQGNPSLSSIRKLIRDG 51
|||||
Db 1380 QLNHEGFKLFATEATSDWLNANNVPANVPANVPAMPVPSQGNPSLSSIRKLIRDG 1430

RESULT 3
US-09-323-472A-12
; Sequence 12, Application US/09323472A
; Patent No. 6346382
; GENERAL INFORMATION:
; APPLICANT: Summar, Marshall
; APPLICANT: Christman, Brian
; TITLE OF INVENTION: HUMAN CARBAMYL PHOSPHATE SYNTHETASE I POLYMORPHISM AND DIAGNOSTIC
; FILE REFERENCE: 1242/19
; CURRENT APPLICATION NUMBER: US/09/323,472A
; CURRENT FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 12
; LENGTH: 1500
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-323-472A-12

Query Match 98.2%; Score 266; DB 4; Length 1500;
Best Local Similarity 98.0%; Pred. No. 2.5e-26;
Matches 50; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 QLNHEGFKLFATEATSDWLNANNVPATPVAMPVPSQGNPSLSSIRKLIRDG 51
|||||
Db 1380 QLNHEGFKLFATEATSDWLNANNVPANVPANVPAMPVPSQGNPSLSSIRKLIRDG 1430

RESULT 4
US-09-323-472A-14
; Sequence 14, Application US/09323472A
; Patent No. 6346382
; GENERAL INFORMATION:
; APPLICANT: Summar, Marshall
; APPLICANT: Christman, Brian
; TITLE OF INVENTION: HUMAN CARBAMYL PHOSPHATE SYNTHETASE I POLYMORPHISM AND DIAGNOSTIC
; FILE REFERENCE: 1242/19
; CURRENT APPLICATION NUMBER: US/09/323,472A
; CURRENT FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 14
; LENGTH: 1500
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-323-472A-14

Query Match 98.2%; Score 266; DB 4; Length 1500;
Best Local Similarity 98.0%; Pred. No. 2.5e-26;
Matches 50; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 QLNHEGFKLFATEATSDWLNANNVPATPVAMPVPSQGNPSLSSIRKLIRDG 51
|||||
Db 1380 QLNHEGFKLFATEATSDWLNANNVPANVPANVPAMPVPSQGNPSLSSIRKLIRDG 1430

RESULT 5
US-09-252-991A-27544
; Sequence 27544, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
```

```
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27544
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27544

Query Match 25.6%; Score 69.5; DB 4; Length 349;
Best Local Similarity 37.1%; Pred. No. 0.35;
Matches 13; Conservative 7; Mismatches 14; Indels 1; Gaps 1;

QY 1 QLNHEGFKLFATEATSDWLNANNVPATPVAMPVPSQ 34
:::|||||
Db 66 RVHSEGEVFLQAWODWLTATGRSHDMLPIAWETR 100

RESULT 6
US-09-252-991A-27861
; Sequence 27861, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27861
; LENGTH: 96
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27861

Query Match 23.2%; Score 63; DB 4; Length 96;
Best Local Similarity 43.8%; Pred. No. 0.49;
Matches 14; Conservative 3; Mismatches 15; Indels 0; Gaps 0;

QY 14 ATSDWLNANNVPATPVAMPVPSQGNPSLSIR 45
:::|||||
Db 47 SSSWPSATRWPAITSVANTSPACIRPSCSYR 78

RESULT 7
US-09-134-001C-4694
; Sequence 4694, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: GPC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4694
; LENGTH: 228
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2003, 02:12:52 ; Search time 19 Seconds
(without alignments)
113.571 Million cell updates/sec

Title: US-09-585-077C-4_COPY_1380_1430
Perfect score: 271
Sequence: 1 QLNHEGFKLFATEATSDWLN.....PSQEGNPSLSIRKLIRDG 51

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS.COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	271	100.0	1500	4	US-09-323-472A-4
2	266	98.2	1500	4	US-09-323-472A-2
3	266	98.2	1500	4	US-09-323-472A-12
4	266	98.2	1500	4	US-09-323-472A-14
5	69.5	25.6	349	4	US-09-252-991A-27544
6	63	23.2	96	4	US-09-252-991A-27861
7	62	22.9	228	4	US-09-134-001C-4694
8	62	22.9	228	4	US-09-134-001C-5495
9	59	21.8	1012	4	US-09-285-385C-4
10	57	21.0	1113	3	US-09-629-616-3
11	55	20.3	167	4	US-09-252-991A-20252
12	55	20.3	441	4	US-08-630-915A-34
13	54	19.9	569	4	US-09-107-532A-6689
14	54	19.9	1073	3	US-09-541-782-6
15	54	19.9	1073	3	US-09-723-820-6
16	54	19.9	3739	3	US-09-320-878-2
17	54	19.9	3739	3	US-09-105-537-3
18	54	19.9	3739	4	US-09-141-908-3
19	54	19.9	3739	4	US-09-657-440-2
20	54	19.9	11877	3	US-09-103-537-6
21	53.5	19.7	878	4	US-09-403-618A-8
22	53	19.6	141	4	US-09-634-238-290
23	53	19.6	153	4	US-09-252-991A-18181
24	53	19.6	227	6	5498499-2
25	53	19.6	276	3	US-08-935-283-16
26	53	19.6	276	4	US-09-594-185-16
27	53	19.6	354	4	US-09-198-452A-536

28 53 19.6 544 2 US-08-932-376A-4 Sequence 4, Appli
29 53 19.6 572 2 US-08-932-376A-2 Sequence 2, Appli
30 52.5 19.4 158 4 US-09-252-991A-30316 Sequence 30316, A
31 52.5 19.4 210 4 US-09-252-991A-25179 Sequence 25179, A
32 52.5 19.4 329 4 US-09-252-991A-30436 Sequence 30436, A
33 52.5 19.4 1708 4 US-09-462-606-2 Sequence 2, Appli
34 52 19.2 364 4 US-08-706-945D-142 Sequence 142, App
35 52 19.2 401 3 US-08-974-022-6 Sequence 6, Appli
36 52 19.2 401 3 US-09-042-785A-12 Sequence 12, Appli
37 52 19.2 401 3 US-08-795-445A-6 Sequence 6, Appli
38 52 19.2 401 3 US-08-795-447A-6 Sequence 6, Appli
39 52 19.2 401 3 US-08-974-186-6 Sequence 6, Appli
40 52 19.2 401 3 US-08-795-446B-6 Sequence 6, Appli
41 52 19.2 401 4 US-08-706-945D-128 Sequence 128, App
42 52 19.2 874 3 US-08-804-439A-15 Sequence 15, Appl
43 52 19.2 874 3 US-08-720-229-15 Sequence 15, Appl
44 52 19.2 1601 4 US-09-345-473E-40 Sequence 40, Appl
45 51.5 19.0 196 4 US-09-252-991A-19503 Sequence 19503, A

ALIGNMENTS

RESULT 1
US-09-323-472A-4
; Sequence 4, Application US/09323472A
; Patent No. 6346382
; GENERAL INFORMATION:
; APPLICANT: Summar, Marshall
; APPLICANT: Christman, Brian
; TITLE OF INVENTION: HUMAN CARBAMYL PHOSPHATE SYNTHETASE I POLYMORPHISM AND DIAGN
; TITLE OF INVENTION: THERAPEUTIC METHODS RELATING THERETO
; FILE REFERENCE: 1242/19
; CURRENT APPLICATION NUMBER: US/09/323,472A
; CURRENT FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 1500
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-323-472A-4

Query Match 100.0%; Score 271; DB 4; Length 1500;
Best Local Similarity 100.0%; Pred. No. 5 5e+27; Indels 0; Gaps 0;
Matches 51; Conservative 0; Mismatches 0;

QY 1 QLNHEGFKLFATEATSDWLNANNVPATVPVWPQEGNPSLSIRKLIRDG 51
|||||
Db 1380 QLNHEGFKLFATEATSDWLNANNVPATVPVWPQEGNPSLSIRKLIRDG 1430

RESULT 2
US-09-323-472A-2
; Sequence 2, Application US/09323472A
; Patent No. 6346382
; GENERAL INFORMATION:
; APPLICANT: Summar, Marshall
; APPLICANT: Christman, Brian
; TITLE OF INVENTION: HUMAN CARBAMYL PHOSPHATE SYNTHETASE I POLYMORPHISM AND DIAGN
; TITLE OF INVENTION: THERAPEUTIC METHODS RELATING THERETO
; FILE REFERENCE: 1242/19
; CURRENT APPLICATION NUMBER: US/09/323,472A
; CURRENT FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 1500
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-323-472A-2

Query Match 98.2%; Score 266; DB 4; Length 1500;


```
Q9ESY1
ID Q9ESY1 PRELIMINARY; PRT; 1803 AA.
AC Q9ESY1;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE TEMO.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Mruk D., Mo M.-Y., Cheng C.-Y.;
RT "TMO is a marker to study sertoli-germ cell interactions: cloning and
RT regulation of a novel testicular molecule.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF296169; AGO9060.1; -
SQ SEQUENCE 1803 AA; 206588 MW; ED73B36E13F25B2 CRC64;

Query Match 23.4%; Score 63.5; DB 11; Length 1803;
Best Local Similarity 37.8%; Pred. No. 62;
Matches 14; Conservative 7; Mismatches 9; Indels 7; Gaps 2;

QY 1 QLNHEGKLFATSDWLNANNVPATPVAMPQEQ 37
Db :||:||||: | |:: |:: | ||
185 ELHKGKLFDELIGLWVQNLP---QW---EQ 214

RESULT 15
Q9N973
ID Q9N973 PRELIMINARY; PRT; 2921 AA.
AC Q9N973;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical 303.3 kDa protein.
GN P1295.15
OS Leishmania major.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Friedlin;
RA Masuy D., Fumelle B., Goffeau A., Ivens A.C., Quail M.,
RA Rajandream M.A., Barrell B.G.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Friedlin;
RX MEDLINE=98146435; PubMed=9477341;
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome.";
RL Genome Res. 8:135-145(1998).
DR EMBL: AL359773; CAB95223.1; -
DR InterPro: IPR000183; Decarboxylase2.
DR PROSITE: PS00879; ODR_DC_2; 1.
KW Hypothetical protein.
SQ SEQUENCE 2921 AA; 303327 MW; FD6B51F33AE37470 CRC64;

Query Match 23.4%; Score 63.5; DB 5; Length 2921;
Best Local Similarity 34.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 10; Mismatches 20; Indels 1; Gaps 1;

QY 4 NEGRKLFATSDWLNANNVPATPVAMPQEQGPSLSIRKLIRD 50
Db :||: | |:: |:: | || | ||: ||: |
2277 NDGELLRTSDSDRSARRV-STETPPQPPQPPSGASLQFTLAD 2322

Search completed: September 13, 2003, 02:15:45
Job time : 60.5 secs
```

Db 174 LSNEGFAIFAGTAVND---TSSVTATATADALHHSNTLDGATIVKLT 219

RESULT 11

Q8K7M5 PRELIMINARY; PRT; 339 AA.
 AC Q8K7M5; MEDLINE=21608550; PubMed=11743193;
 DT 01-JUN-2002 (TRENBLrel. 22, Created)
 DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
 DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
 DE Hypothetical protein SppM3_0740.
 GN SPM3_0740.
 OS Streptococcus pyogenes (serotype M3).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=198466;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MGAS315 / Serotype M3;
 RX MEDLINE=22133808; PubMed=12122206;
 RA Beres S.B., Sylva G.L., Barbican K.D., Lei B., Hoff J.S.,
 RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,
 RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
 RA Schlievert P.M., Musser J.M.;
 RT "Genome sequence of a serotype M3 strain of group A Streptococcus;
 RT phage-encoded toxins, the high-virulence phenotype, and clone
 RT emergence";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).
 DR EMBL; AB014151; AAM79347.1;
 DR InterPro: IPR003439; ABC transporter.
 DR InterPro: IPR004630; Cons_hypoth698.
 DR Pfam; PF03601; Cons_hypoth698; 1.
 DR PROSITE; PS00211; ABC-TRANSPORTER; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 339 AA; 36275 MW; 4C1D2C9A3C94283B CRC64;

Query Match 23.6%; Score 64; DB 16; Length 339;
 Best Local Similarity 38.8%; Pred. No. 8;
 Matches 19; Conservative 7; Mismatches 19; Indels 4; Gaps 2;

Qy 2 LHNEGKLFATEATSDWLNANNVPATVPVWPSQGNP-SLSIRKLIR 49
 Db 174 LSNEGFAIFAGTAVND---TSSVTATATADALHHSNTLDGATIVKLT 219

RESULT 12

Q8UG58 PRELIMINARY; PRT; 863 AA.
 AC Q8UG58; MEDLINE=21608550; PubMed=11743193;
 DT 01-JUN-2002 (TRENBLrel. 21, Created)
 DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
 DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
 DE DNA-directed RNA polymerase.
 GN ATU1183 OR AGR_C_2186.
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Rhizobium.
 OX NCBI_TaxID=176299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21608550; PubMed=11743193;
 RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
 RA Okura Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
 RA Chen V., Paulsen I.T., Eisen J.A., Karp P.D., Boree D. Sr.,
 RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
 RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
 RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
 RA Zhang S., Tao H., Biddle P., Jung M., Krespan W., Perry M.,
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
 RA Nester E.W.;
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens

C58.";
 RL Science 294:2317-2323(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21608551; PubMed=11743194;
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
 RA Quirio B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
 RA Houmlel K., Gordon J., Vaudin M., Harchouk O., Epp A., Liu F.,
 RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
 RA Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
 RA Cielo C., Slater S.;
 RT "Genome sequence of the plant pathogen and biotechnology agent
 RT Agrobacterium tumefaciens C58.";
 RL Science 294:2323-2328(2001).
 DR EMBL; AB009081; AAL42195.1;
 DR EMBL; AB008047; AAK86987.1;
 DR InterPro: IPR002092; RNA_pol_phage.
 DR Pfam; PF00940; RNA_pol; 1.
 DR PROSITE; PS00900; RNA_POL_PHAGE_1; 1.
 DR PROSITE; PS00489; RNA_POL_PHAGE_2; 1.
 KW DNA-directed RNA polymerase; Complete proteome.
 SQ SEQUENCE 863 AA; 97075 MW; 9B038C4FB44D36CA CRC64;

Query Match 23.6%; Score 64; DB 16; Length 863;
 Best Local Similarity 30.2%; Pred. No. 23;
 Matches 19; Conservative 7; Mismatches 23; Indels 14; Gaps 2;

Qy 3 HNEGKLFATEATSDWLNANNVPATVPVWPSQGNP-SLSIR-----KLI 48
 Db 221 HNTIYVATKTDWLATENSRLAPLSPLYLTLPVPPRWTSPFRGGYWSGRVRLRI 280

Qy 49 RDG 51

Db 281 KTG 283

RESULT 13

Q14997 PRELIMINARY; PRT; 1798 AA.
 AC Q14997; MEDLINE=96051398; PubMed=7584044;
 DT 01-NOV-1996 (TRENBLrel. 01, Created)
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
 DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
 DE Hypothetical protein KIAA0077 (Fragment).
 GN KIAA0077.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96051398; PubMed=7584044;
 RA Nomura N., Nagase T., Miyajima N., Sazuka T., Tanaka A., Sato S.,
 RA Seki N., Kawabayashi Y., Ishikawa K., Tabata S.;
 RT "Prediction of the coding sequences of unidentified human genes. II.
 RT The coding sequences of 40 new genes (KIAA0041-KIAA0080) deduced by
 RT analysis of cDNA clones from human cell line KG-1.";
 RL DNA Res. 1:223-229(1994).
 DR EMBL; D38521; BAA07526.1;
 KW Hypothetical protein.
 FT NON_TER 1
 SQ SEQUENCE 1798 AA; 206406 MW; 6FCB2F2AA77999C1 CRC64;

Query Match 23.4%; Score 63.5; DB 4; Length 1798;
 Best Local Similarity 37.8%; Pred. No. 62;
 Matches 14; Conservative 7; Mismatches 9; Indels 7; Gaps 2;

Qy 1 QHNEGKLFATEATSDWLNANNVPATVPVWPSQEQ 37

Db 180 ELHHKGFKEFDELGLVSVQNLPL---QW---EQ 209

RESULT 14

DR InterPro; IPR001317; CPS_Gatase.
DR InterPro; IPR000991; Gatase_1.
DR InterPro; IPR004362; MGS_like.
DR InterPro; IPR006131; OTCace_O.
DR InterPro; IPR006132; OTCace_P.
DR Pfam; PF00289; CPsase_L_chain; 2.
DR Pfam; PF02786; CPsase_L_D2; 3.
DR Pfam; PF02787; CPsase_L_D3; 1.
DR Pfam; PF00988; CPsase_sm_chain; 1.
DR Pfam; PF00117; Gatase; 1.
DR Pfam; PF02142; MGS; 1.
DR Pfam; PF00185; OTCace; 1.
DR Pfam; PF02729; OTCace_N; 1.
DR PRINTS; PRO0097; ANTSNTHASEII.
DR PRINTS; PRO0100; AOTCASE.
DR PRINTS; PRO0098; CPsase.
DR PRINTS; PRO0099; CPsGATASE.
DR PRINTS; PRO0096; GATASE.
DR TIGRFAMs; TIGR00670; asp_carb_tr; 1.
DR TIGRFAMs; TIGR01368; CPsaseIISmall; 1.
DR TIGRFAMs; TIGR01369; CPsaseII_lig; 1.
DR PROSITE; PS00097; CARBAMOYLTRANSFERASE; 1.
DR PROSITE; PS00866; CPsase_I; 2.
DR PROSITE; PS00867; CPsase_2; 2.
DR PROSITE; PS00442; GATASE_TYPE_I; 1.
KW Ligase.
SQ SEQUENCE 2275 AA; 249739 MW; 27BA9C1FA751436A CRC64;

Query Match 25.5%; Score 69; DB 3; Length 2275;
Best local Similarity 52.0%; Pred.No.16;
Matches 13; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 QLHNEGFKLFATEATSDMLNANNVP 25
 | : | : | : | : | : | : | : | : |
DB 1422 KLRDVGFLFATSGTADFLKENGVP 1446

RESULT 10
Q99ZV5
ID Q99ZV5 PRELIMINARY; PRT; 339 AA.
AC Q99ZV5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein Spv1056.
GN SPV1056.
GE Streptococcus pyogenes.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
ON NCBI_taxID=1314;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SF370 / ATCC 700294 / Serotype M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferritin J.J., McShan W.M., Ajdic D.J., Savic D.J., Savić G., Lyon K.,
RA Perreault C., Sezate S., Suvarov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najaf F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL; AE006550; AAK33943.1; -.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR004630; Cons_hypoth698.
DR Pfam; PF03601; Cons_hypoth698; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 339 AA; 36239 MW; A03E1866CF0DD99A CRC64;

```

QY      1  QLHNEGKLFATEATSDWLNNANNPATPVAWPSQEGQNPSLSIRKLIRDG 51
      III : : : : : III : : : : : III : : : : : I : I : : : :
DB      1381  QLHQEGYKLYATEGTSTLWLVNDVPTAPYSWPTAEDHSSAPSTFKLIHDG 1431

RESULT 6
Q9PTM9      PRELIMINARY;      PRT; 1504 AA.
ID      Q9PTM9
AC      Q9PTM9;
DT      01-MAY-2000 (TrEMBLrel. 13, Created)
DT      01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT      01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE      Aciyglutamate-activated carbamoyl phosphate synthase III.
DE      Opsanus beta.
OS      Opsanus beta.
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC      Acanthomorpha; Paracanthopterygii; Batrachoididae; Opsanus.
OX      NCBI_taxid=95145;
RN      [1]
RS      SEQUENCE FROM N.A.
RC      TISSUE=Liver;
RC      Kong H., Kalatapitiya N., Kingsley K., Salo W.L., Anderson P.M.,
RA      Wang Y.S., Walsh P.J.;
RA      "Induction of expression of carbamoyl phosphate synthetase III and
RT      glutamine synthetase mRNA during confinement stress in gulf toadfish
RT      (Opsanus beta).";
RL      Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AF169248; RAD51318.2; -.
DR      HSSP; P00968; IA9X.
DR      InterPro; IPR006275; CarA_L_glu.
DR      InterPro; IPR006274; CarA_small.
DR      InterPro; IPR005483; Cpase_L.
DR      InterPro; IPR005479; Cpase_L_D2.
DR      InterPro; IPR005480; Cpase_L_D3.
DR      InterPro; IPR005481; Cpase_L_N.
DR      InterPro; IPR002474; CPSase_sm_chain.
DR      InterPro; IPR001317; CPS_GATase.
DR      InterPro; IPR000991; Gatase_1.
DR      InterPro; IPR004362; MGS_like.
DR      Pfam; PF00289; CPSase_L_chain; 2.
DR      Pfam; PF02786; CPSase_L_D2; 2.
DR      Pfam; PF02787; CPSase_L_D3; 1.
DR      Pfam; PF00988; CPSase_sm_chain; 1.
DR      Pfam; PF00117; Gatase; 1.
DR      Pfam; PF02142; MGS; 1.
DR      PRINTS; PR00096; CPSASE.
DR      PRINTS; PR00099; CPSGATASE.
DR      PRINTS; PR00096; GATASE.
DR      TIGRfams; TIGR01368; CPSaseIIsmall; 1.
DR      TIGRfams; TIGR01369; CPSaseII_lrg; 1.
DR      PROSITE; PS00866; CPSASE_1; 2.
DR      PROSITE; PS00867; CPSASE_2; 4.
SQ      SEQUENCE 1504 AA; 165446 MW; C46D618F65AA3F8D CRC64;

Query Match      59.6%; Score 161.5; DB 13; Length 1504;
Best Local Similarity 62.7%; Pred. No. 2.1e-11;
Matches 32; Conservative 10; Mismatches 8; Indels 1; Gaps

QY      1  QLHNEGKLFATEATSDWLNNANNPATPVAWPSQEGQNPSLSIRKLIRDG 51
      III : : : : : III : : : : : III : : : : : I : I : : : :
DB      1380  QLKDEGKLYATEATSNLCANDVSPVAVWFS-DNADSNLPKIRLISEG 1429

RESULT 7
Q18990      PRELIMINARY;      PRT; 2198 AA.
ID      Q18990
AC      Q18990;
DT      01-NOV-1996 (TrEMBLrel. 01, Created)
DT      01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DE      01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE      D2085.1 protein.
DE      D2085.1.
GN      D2085.1.

```

OS	Caeenorhabditis elegans.
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;
OT	Rhabditiidae; Peloderrinae; Caeenorhabditis.
OX	NCBI_TaxID=6239;
[1]	
RN	SEQUENCE FROM N.A.
RP	Raynes C.;
RL	Submitted (OCT-1995) to the EMBL/GenBank/DDBJ databases.
RA	[2]
RN	SEQUENCE FROM N.A.
RP	
RX	MEDLINE=99069613; PubMed=9851916;
RA	none;
RT	"Genome sequence of the nematode C.elegans: A platform for
RT	investigating biology.";
RT	Science 282:2012-2018(1998).
DR	EMBL; Z54284; CAA91059.1; -;
DR	HSSP; P00479; 3CSU.
DR	WormPep; D2085.1; CEQ3105.
DR	InterPro; IPR006680; Amidohydro.L.
DR	InterPro; IPR006220; Anth_synthII.
DR	InterPro; IPR006130; Asp/Orn_Cotranf.
DR	InterPro; IPR002082; Asp_carbmItransf.
DR	InterPro; IPR006275; CARA_L_glu.
DR	InterPro; IPR006274; CARA_small.
DR	InterPro; IPR005483; Cpase_I.
DR	InterPro; IPR005479; Cpase_I_D2.
DR	InterPro; IPR005480; Cpase_I_D3.
DR	InterPro; IPR005481; Cpase_I_N.
DR	InterPro; IPR002474; CPSase_sm_chain.
DR	InterPro; IPR001317; CPS_Garase.
DR	InterPro; IPR002195; Dihydroarase.
DR	InterPro; IPR000991; Gatrase_1.
DR	InterPro; IPR004362; MGS_like.
DR	InterPro; IPR006131; OTCace_O.
DR	InterPro; IPR006132; OTCace_P.
DR	InterPro; IPR005847; Urease.
DR	Pfam; PF01979; Amidohydro_L; 1.
DR	Pfam; PF00289; CPSase_L_chain; 2.
DR	Pfam; PF02786; CPSase_L_D2; 2.
DR	Pfam; PF02787; CPSase_L_D3; 1.
DR	Pfam; PF00988; CPSase_sm_chain; 1.
DR	Pfam; PF00117; Gatase; 1.
DR	Pfam; PF02142; MGS; 1.
DR	Pfam; PF00185; OTCace; 1.
DR	Pfam; PF02729; OTCace_N; 1.
DR	PRINTS; PR00097; ANTSENTHASEII.
DR	PRINTS; PR00100; AOTCACE.
DR	PRINTS; PR00098; CPSASE.
DR	PRINTS; PR00099; CPSCGATASE.
DR	PRINTS; PR00096; GATASE.
DR	Prodom; PD000518; Urease; 1.
DR	TIGFAMS; TIGR00670; asp_cab.tr; 1.
DR	TIGFAMS; TIGR01368; CPSaseIISmall; 1.
DR	TIGFAMS; TIGR01369; CPSaseII_lrig; 1.
DR	PROSITE; PS00097; CARBAMOYLTRANSFERASE; 1.
DR	PROSITE; PS00866; CPSASE_1; 2.
DR	PROSITE; PS00867; CPSASE_2; 2.
DR	PROSITE; PS00482; DIHYDROOROTASE_1; 1.
DR	PROSITE; PS00483; DIHYDROOROTASE_2; 1.
DR	PROSITE; PS00442; GATASE TYPE_I; 1.
SQ	SEQUENCE 2198 AA; 242567 MW; 2C029AFD34C71A4E CRC64;
Query Match	29.2%; Score 79; dB 5; Length 2198;
Best Local Similarity	33.3%; Pred.No. 0.86;
Matches	11; Conservative 11; Mismatches 11; Indels 0; Gaps 0;
QY	6 GFKLFEATEATSQWLNNANVPATPVAMPVSEGGN 38
Db	1349 GYEYSGKTADYFQSINKINVKPDWFEEGSS 1381
RESULT 8	
Q916H0	

DR InterPro: IPR000991; GATase.1.
 DR InterPro: IPR004362; MGS_like.
 DR Pfam: PF00289; CPSase_L_chain; 2.
 DR Pfam: PF02786; CPSase_L_D2; 2.
 DR Pfam: PF02787; CPSase_L_D3; 1.
 DR Pfam: PF00988; CPSase_sm_chain; 1.
 DR Pfam: PF00117; GATase; 1.
 DR Pfam: PF02142; MGS; 1.
 DR PRINTS: PR00098; CPSASE.
 DR PRINTS: PR00099; CPSGATASE.
 DR TIGRFAMS: TIGR01368; CPSaseIsmall; 1.
 DR TIGRFAMS: TIGR01369; CPSaseII_lrg; 1.
 DR PROSITE: PS00866; CPSASE.1; 2.
 DR PROSITE: PS00867; CPSASE.2; 2.
 DR SEQUENCE 1505 AA; 164592 MW; 89CA422285166CA2 CRC64;
 Query Match 61.3%; Score 166; DB 13; Length 1505;
 Best Local Similarity 64.7%; Pred. No. 5.5e-12;
 Matches 33; Conservative 7; Mismatches 11; Indels 0; Gaps 0;
 QY 1 QLNHEGKFLFATEASDLNANNVPATVPVWPSQEQNPGLSIRKLFDG 51
 || ||||| ||||| || ||||| ||||| : : ||||| : :
 Db 1380 QLKKEGKFLYATEASDLNANNVPATVPVWPTGMDGNSLPSIRLISEG 1430
 RESULT 5
 Q92115
 ID Q92115 PRELIMINARY; PRT; 1502 AA.
 AC Q92115; Q91438;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-NOV-2003 (TrEMBLrel. 23, Last annotation update)
 DE Carbamoyl phosphate synthetase III [glutamine] mitochondrial precursor
 DE (EC 6.3.5.5) (Glutamine-dependent carbamoyl-phosphate synthetase III)
 DE (GD-CPSase III).
 GN CPSIII.
 OS Squalus acanthias (Spiny dogfish).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
 OC Elasmobranchii; Squala; Squaloidei; Squalidae; Squalus.
 OX NCBI_TaxID-7797;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 39-47.
 RC TISSUE=LIVER;
 RX MEDLINE=95018247; PubMed=7932737;
 RA Hong J., Salo W.L., Lusty C.J., Anderson P.M.;
 RT "Carbanyl phosphate synthetase III, an evolutionary intermediate in
 RT the transition between glutamine-dependent and ammonia-dependent
 RT carbamyl phosphate synthetases.";
 RL J. Mol. Biol. 243:131-140(1994).
 RN [2]
 RP SEQUENCE OF 1-42 FROM N.A.
 RC TISSUE=TESTIS;
 RX MEDLINE=97143029; PubMed=8995057;
 RA Hong J., Salo W.L., Chen Y., Atkinson B.G., Anderson P.M.;
 RT "The promoter region of the carbamoyl-phosphate synthetase III gene of
 RT Squalus acanthias.";
 RL J. Mol. Biol. Evol. 43:602-609(1996).
 RN [3]
 RP ENZYME ACTIVITY, AND REGULATION.
 RC TISSUE=LIVER;
 RX MEDLINE=80168997; PubMed=6245445;
 RA Anderson P.M.;
 RT "Glutamine- and N-acetylglutamate-dependent carbamoyl phosphate
 RT synthetase in elasmobranchs.";
 RL Science 208:291-293(1980).
 RN [4]
 RP ENZYME ACTIVITY, AND SUBUNITS.
 RC TISSUE=LIVER;
 RX MEDLINE=82053090; PubMed=7298655;
 RA Anderson P.M.;
 RT "Purification and properties of the glutamine- and N-acetyl-L-
 RT glutamate-dependent carbamoyl phosphate synthetase from liver of
 RT Squalus acanthias.";

RL J. Biol. Chem. 256:12228-12238(1981).
 RN [5]
 RP SUBCELLULAR LOCATION.
 RC TISSUE=LIVER;
 RX MEDLINE=82214067; PubMed=6123510;
 RA Casey C.A., Anderson P.M.;
 RT "Subcellular location of glutamine synthetase and urea cycle enzymes
 RT in liver of spiny dogfish (Squalus acanthias).";
 RL J. Biol. Chem. 257:8449-8453(1982).
 CC -!- FUNCTION: INVOLVED IN THE UREA CYCLE OF UREOTELIC ANIMALS WHERE
 CC THE ENZYME PLAYS AN IMPORTANT ROLE IN THE GLUTAMINE-DEPENDENT
 CC FORMATION OF CARBAMOYL PHOSPHATE FOR UREA SYNTHESIS. UREA MAYBE
 CC RETAINED IN TISSUES AND BODY FLUIDS AS A MECHANISM OF
 CC OSMOREGULATION.
 CC -!- CATALYTIC ACTIVITY: 2 ATP + GLUTAMINE + CO(2) + H(2)O -> 2 ADP +
 CC PHOSPHATE + GLUTAMATE + CARBAMOYL PHOSPHATE.
 CC -!- ENZYME REGULATION: REQUIRES N-ACETYLGLUTAMATE AS AN ALLOSTERIC
 CC ACTIVATOR.
 CC -!- SUBUNIT: MONOMER.
 CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
 CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN LIVER. MAY NOT BE
 CC EXPRESSED IN OTHER TISSUES.
 CC -!- SIMILARITY: TO OTHER CARBAMOYL-PHOSPHATE SYNTHETASES. ALSO
 CC CONTAINS A GLUTAMINE AMIDOTRANSFERASE-LIKE DOMAIN. THE CYSTEINE
 CC RESIDUE WHICH IS ESSENTIAL FOR AMIDO TRANSFERASE ACTIVITY, IS
 CC CONSERVED.
 DR EMBL: L31362; AAA96435.1; -;
 DR EMBL: U19771; AAB49032.1; -;
 DR HSSP: P00968; LCSO.
 DR InterPro: IPR006275; CarA_L_glu.
 DR InterPro: IPR006274; CarA_small.
 DR InterPro: IPR005483; CPase_L.
 DR InterPro: IPR005479; CPase_L_D2.
 DR InterPro: IPR005480; CPase_L_D3.
 DR InterPro: IPR005481; CPase_L_N.
 DR InterPro: IPR002474; CPSase_sm_chain.
 DR InterPro: IPR001317; CPS_GATase.
 DR InterPro: IPR000991; GATase.1.
 DR InterPro: IPR004362; MGS_like.
 DR Pfam: PF00289; CPSase_L_chain; 2.
 DR Pfam: PF02786; CPSase_L_D2; 2.
 DR Pfam: PF02787; CPSase_L_D3; 1.
 DR Pfam: PF00988; CPSase_sm_chain; 1.
 DR Pfam: PF00117; GATase; 1.
 DR Pfam: PF02142; MGS; 1.
 DR PRINTS: PR00098; CPSASE.
 DR PRINTS: PR00099; CPSGATASE.
 DR PRINTS: PR00096; GATASE.
 DR TIGRFAMS: TIGR01368; CPSaseIsmall; 1.
 DR TIGRFAMS: TIGR01369; CPSaseII_lrg; 1.
 DR PROSITE: PS00866; CPSASE.1; 2.
 DR PROSITE: PS00867; CPSASE.2; 2.
 KW Ligase; Transil peptide; Mitochondrion; ATP-binding; Urea cycle.
 FT CHAIN 1 38 MITOCHONDRION.
 FT CHAIN 39 1502 [GLUTAMINE] MITOCHONDRIAL
 FT DOMAIN 39 219 ANTHRANILATE PHOSPHORIBOSYLTRANSFERASE-
 FT LIKE.
 FT DOMAIN 220 411 GLUTAMINE AMIDOTRANSFERASE-LIKE.
 FT DOMAIN 412 1502 CARBAMOYL-PHOSPHATE SYNTHETASE LARGE
 FT CHAIN.
 FT NP_BIND 572 627 ATP (2) (POTENTIAL).
 FT NP_BIND 719 769 ATP (1) (POTENTIAL).
 FT NP_BIND 1114 1172 ATP (2) (POTENTIAL).
 FT NP_BIND 1259 1303 ATP (1) (POTENTIAL).
 FT REPEAT 420 877
 FT REPEAT 971 1411
 FT ACT_SITE 294 294 GATASE (BY SIMILARITY).
 SQ SEQUENCE 1502 AA; 165140 MW; CF44C222D99E2EE CRC64;

Query Match 60.9%; Score 165; DB 13; Length 1502;
 Best Local Similarity 56.9%; Pred. No. 7.4e-12;
 Matches 29; Conservative 7; Mismatches 15; Indels 0; Gaps 0;

```

DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Carbamoyl-phosphate synthetase III (EC 6.3.5.5).
OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX NCBI_TaxID=8022;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97197794; PubMed=9045644;
RA Korte J.J., Salo W.L., Cabrera V.M., Wright P.A., Feltskie A.K.,
RA Anderson P.M.;
RT "Expression of carbamoyl-phosphate synthetase III mRNA during the
RT early stages of development and in muscle of adult rainbow trout
RT (Oncorhynchus mykiss).";
RL J. Biol. Chem. 272:6270-6277(1997).
DR EMBL; U65893; AAC60207.1; -
DR HSSP; P00968; 1CS0.
DR InterPro; IPR006275; CarA_L_glu.
DR InterPro; IPR006274; CarA_small.
DR InterPro; IPR005483; CPase_L.
DR InterPro; IPR005479; CPase_L_D2.
DR InterPro; IPR005473; CPase_L_D3.
DR InterPro; IPR005480; CPase_L_N.
DR InterPro; IPR002474; CPase_sm_chain.
DR InterPro; IPR001317; CPS_GATase.
DR InterPro; IPR000991; GATase_1.
DR InterPro; IPR004362; MGS_like.
DR Pfam; PF02786; CPase_L_D2; 2.
DR Pfam; PF02787; CPase_L_D3; 1.
DR Pfam; PF00988; CPase_sm_chain; 1.
DR Pfam; PF00117; GATase; 1.
DR Pfam; PF02142; MGS; 1.
DR PRINTS; PR00098; CPASE.
DR PRINTS; PR00096; GATASE.
DR TIGRfams; TIGR01368; CPaseIIsmall; 1.
DR TIGRfams; TIGR01369; CPaseII_lrg; 1.
DR PROSITE; PS00866; CPASE_1; 1.
DR PROSITE; PS00867; CPASE_2; 2.
KW Ligase.
SQ SEQUENCE 1518 AA; 166577 MW; F9D6331C170A0EBD CRC64;

Query Match 72.0%; Score 195; DB 13; Length 1518;
Best Local Similarity 74.5%; Pred. No. 1.2e-15;
Matches 38; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

QY 1 QLNHEGFKLPATATSDMLNANNVPATVPVWPSQEGONPSLSIRKLIRDG 51
DB 1378 QLNHEGFKLPATATSDMLNANNVPATVPVWPSQEGONPSLSIRKLIRDG 1428

RESULT 3
O42433
ID O42433 PRELIMINARY; PRT; 1506 AA.
AC O42433
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Carbamoyl-phosphate synthetase III (EC 6.3.5.5).
OS Micropterus salmoides (Largemouth bass).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Perciformes; Percoidae;
OC Centrarchidae; Micropterus.
OX NCBI_TaxID=27706;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Muscle;
RA Kong H., Edberg D.D., Salo W.L., Korte J.J., Wright P.A.,
RA Anderson P.M.;

```

```

RT "Nitrogen Excretion and Expression of Carbamoyl-Phosphate Synthetase
RT III in Extra-Hepatic Tissues of Largemouth Bass (Micropterus
RT salmoides).";
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF006491; AAB62566.1; -
DR HSSP; P00968; 1CS0.
DR InterPro; IPR006275; CarA_L_glu.
DR InterPro; IPR006274; CarA_small.
DR InterPro; IPR005483; CPase_L.
DR InterPro; IPR005479; CPase_L_D2.
DR InterPro; IPR005473; CPase_L_D3.
DR InterPro; IPR005480; CPase_L_N.
DR InterPro; IPR002474; CPase_sm_chain.
DR InterPro; IPR001317; CPS_GATase.
DR InterPro; IPR000991; GATase_1.
DR InterPro; IPR004362; MGS_like.
DR Pfam; PF02786; CPase_L_D2; 2.
DR Pfam; PF02787; CPase_L_D3; 1.
DR Pfam; PF00988; CPase_sm_chain; 1.
DR Pfam; PF00117; GATase; 1.
DR Pfam; PF02142; MGS; 1.
DR PRINTS; PR00098; CPASE.
DR PRINTS; PR00096; GATASE.
DR TIGRfams; TIGR01368; CPaseIIsmall; 1.
DR TIGRfams; TIGR01369; CPaseII_lrg; 1.
DR PROSITE; PS00866; CPASE_1; 1.
DR PROSITE; PS00867; CPASE_2; 2.
KW Ligase.
SQ SEQUENCE 1506 AA; 165304 MW; 59439669AC8BB86D CRC64;

Query Match 64.9%; Score 176; DB 13; Length 1506;
Best Local Similarity 66.0%; Pred. No. 3e-13;
Matches 33; Conservative 9; Mismatches 8; Indels 0; Gaps 0;

QY 2 LHNEGFKLPATATSDMLNANNVPATVPVWPSQEGONPSLSIRKLIRDG 51
DB 1382 LHNEGFKLPATATSDMLNANNVPATVPVWPSQEGONPSLSIRKLIRDG 1431

RESULT 4
Q9W7F3
ID Q9W7F3 PRELIMINARY; PRT; 1505 AA.
AC Q9W7F3.
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Carbamoyl-phosphate synthetase III.
OS Alcolapia grahami.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Perciformes; Labroidae;
OC Cichlidae; Alcolapia.
OX NCBI_TaxID=87886;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99445540; PubMed=10514466;
RA Lindsey T.E., Scheiderer C.H., Walsh P.J., Wood C.M., Bergman H.L.,
RA Bergman A.L., Laurent P., Wilson P., Anderson P.M.;
RA "Muscle as the primary site of urea cycle enzyme activity in an
RT alkaline lake-adapted tilapia, Oreochromis alcalicus grahami.";
RL J. Biol. Chem. 274:29858-29861(1999).
DR EMBL; AF119250; AAD43968.1; -
DR HSSP; P00968; 1A9X.
DR InterPro; IPR006275; CarA_L_glu.
DR InterPro; IPR006274; CarA_small.
DR InterPro; IPR005483; CPase_L.
DR InterPro; IPR005479; CPase_L_D2.
DR InterPro; IPR005480; CPase_L_D3.
DR InterPro; IPR005481; CPase_L_N.
DR InterPro; IPR002474; CPase_sm_chain.
DR InterPro; IPR001317; CPS_GATase.

```

GenCore version 5.1.6
Copyright (C) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model
Run on: September 13, 2003, 02:07:12 ; Search time 57.5 Seconds
(without alignments)
228.881 Million cell updates/sec

Title: US-09-585-077c-4_COPY_1380_1430
Perfect score: 271
Sequence: 1 QLNHEGFKLFATSDWLN.....PSQEGNPSLSIRKLRDQ 51

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.rodent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*
15: sp.rvirus:*
16: sp.bacteriap:*
17: sp.archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	271	100.0	748	11 Q8C196	Q8C196 mus musculus
2	195	72.0	1518	13 P70087	P70087 oncorhynch
3	176	64.9	1506	13 Q42433	Q42433 micropterus
4	166	61.3	1505	13 Q9W7F3	Q9W7F3 alcolapia g
5	165	60.9	1502	13 Q92115	Q92115 squalus aca
6	161.5	59.6	1504	13 Q9PTM9	Q9PTM9 opsanus bet
7	79	29.2	2198	5 Q18990	Q18990 caenorhabdi
8	69.5	25.6	344	16 Q916H0	Q916H0 pseudomonas
9	69	25.5	2275	3 Q93937	Q93937 emericeila
10	64	23.6	339	16 Q992V5	Q992V5 streptococc
11	64	23.6	339	16 Q8K7M5	Q8K7M5 streptococc
12	64	23.6	863	16 Q8UG58	Q8UG58 agrobacteri
13	63.5	23.4	1798	4 Q14997	Q14997 homo sapien
14	63.5	23.4	1803	11 Q9ESY1	Q9ESY1 rattus norv
15	63.5	23.4	2921	5 Q9N973	Q9N973 leishmania
16	63	23.2	339	16 Q8P171	Q8P171 streptococc

17	63	23.2	540	3 Q00878	Q00878 nectria hae
18	63	23.2	1065	16 Q8EZX0	Q8EZX0 leptospira
19	63	23.2	2204	12 Q99FK6	Q99FK6 porcine tes
20	62	22.9	336	16 Q97TA9	Q97TA9 streptococ
21	62	22.9	340	16 Q8DRN6	Q8DRN6 streptococ
22	62	22.9	3564	11 Q92313	Q92313 mus musculu
23	61.5	22.7	1068	16 Q8CXH7	Q8CXH7 oceanobacil
24	61	22.5	185	5 Q8IPB6	Q8IPB6 drosophila
25	61	22.5	510	5 Q9VKT5	Q9VKT5 drosophila
26	60.5	22.3	246	4 Q9UJL8	Q9UJL8 homo sapien
27	60.5	22.3	1360	12 Q55253	Q55253 murine hepa
28	60	22.1	157	16 Q8P5C0	Q8P5C0 xanthomonas
29	60	22.1	341	16 Q8DRZ4	Q8DRZ4 streptococ
30	60	22.1	455	16 Q8PJ84	Q8PJ84 xanthomonas
31	60	22.1	889	2 Q8XNY4	Q8XNY4 xanthomonas
32	60	22.1	1113	16 Q8FT42	Q8FT42 corynebacte
33	59.5	22.0	508	4 Q8IZH4	Q8IZH4 homo sapien
34	59.5	22.0	537	16 Q9A8M4	Q9A8M4 caulobacter
35	59.5	22.0	1181	4 Q9UM06	Q9UM06 homo sapien
36	59.5	22.0	1220	4 Q9UM05	Q9UM05 homo sapien
37	59.5	22.0	1220	4 Q94973	Q94973 homo sapien
38	59.5	22.0	1249	11 Q922Q1	Q922Q1 rattus norv
39	59.5	22.0	1698	12 Q9YLR3	Q9YLR3 hepatitis e
40	59	21.8	375	10 Q8VXK9	Q8VXK9 fagus sylv
41	59	21.8	523	16 Q8PKJ9	Q8PKJ9 xanthomonas
42	59	21.8	1012	11 Q9WVM6	Q9WVM6 mus musculu
43	58.5	21.6	277	16 Q92DF9	Q92DF9 listeria in
44	58.5	21.6	277	16 Q8Y8N5	Q8Y8N5 listeria mo
45	58.5	21.6	715	3 Q12204	Q12204 saccharomyc

ALIGNMENTS

RESULT 1

Q8C196 ID Q8C196 PRELIMINARY; PRT; 748 AA.
AC Q8C196;
DT 01-MAR-2003 (TREMREL. 23, Created)
DT 01-MAR-2003 (TREMREL. 23, Last sequence update)
DT 01-MAR-2003 (TREMREL. 23, Last annotation update)
DE Weakly similar to carbamoyl-phosphate synthase (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=SKIN;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK028683; BAC26064.1; -
FT NON_TER
SQ SEQUENCE 748 AA; 83277 MW; AA818F3BAB82D9A1 CRC64;

Query Match 100.0%; Score 271; DB 11; Length 748;
Best Local Similarity 100.0%; Pred. No. 1.3e-25;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QLNHEGFKLFATSDWLNANNVPATVPVWPQEGNPSLSIRKLRDQ 51
|||||
Db 628 QLNHEGFKLFATSDWLNANNVPATVPVWPQEGNPSLSIRKLRDQ 678

RESULT 2

P70087 ID P70087 PRELIMINARY; PRT; 1518 AA.
AC P70087;
DT 01-FEB-1997 (TREMREL. 02, Created)


```

DR PFAM: PF00185; OTCace; 1.
DR PFAM: PF02729; OTCace_N; 1.
DR PRINTS: PR00097; ANTSENTHASEII.
DR PRINTS: PR00100; OTCASE.
DR PRINTS: PR00098; CPSASE.
DR PRINTS: PR00099; CPSGATASE.
DR PRINTS: PR00096; GATASE.
DR PRODOM: PD000518; Urease; 1.
DR TIGRFAMS: TIGR00670; asp_cab.tr; 1.
DR TIGRFAMS: TIGR01369; CPSaseI.lrg; 1.
DR TIGRFAMS: TIGR01368; CPSaseI.small; 1.
DR TIGRFAMS: TIGR00857; PYC.mlti; 1.
DR PROSITE: PS00097; CARBAMOYLTRANSFERASE; 1.
DR PROSITE: PS00442; GATASE.TYPE_I; 1.
DR PROSITE: PS00482; DIHYDROOROTASE_1; 1.
DR PROSITE: PS00483; DIHYDROOROTASE_2; 1.
DR PROSITE: PS00866; CPSASE_1; 2.
DR PROSITE: PS00867; CPSASE_2; 2.
KW Pyrimidine biosynthesis; Ligase; Transferase; Hydrolase; Zinc;
KW Allosteric enzyme; Multifunctional enzyme; Phosphorylation.
FT DOMAIN 1 365 GATASE (GLUTAMINE AMIDOTRANSFERASE).
FT DOMAIN 366 394 LINKER.
FT DOMAIN 395 1455 CPSASE (CARBAMOYL-PHOSPHATE SYNTHASE).
FT DOMAIN 395 933 CPSASE A.
FT DOMAIN 934 1455 CPSASE B.
FT DOMAIN 1456 1788 DHOASE (DIHYDROOROTASE).
FT DOMAIN 1789 1917 LINKER.
FT DOMAIN 1918 2225 ATCASE (ASPARTATE TRANS-CARBAMYLASE).
FT ACT_SITE 252 252 GATASE (BY SIMILARITY).
FT ACT_SITE 336 336 GATASE (BY SIMILARITY).
FT ACT_SITE 338 338 GATASE (BY SIMILARITY).
FT METAL 1471 1471 ZINC (POTENTIAL).
FT METAL 1473 1473 ZINC (POTENTIAL).
SQ SEQUENCE 2225 AA; 243126 MW; 9F6EBA9BD4C6E5A CRC64;

Query Match 22.0%; Score 59.5; DB 1; Length 2225;
Best Local Similarity 32.6%; Pred. No. 50;
Matches 15; Conservative 9; Mismatches 19; Indels 3; Gaps 1;

QY 2 LHNEGFKLFATEATSDWLNANNVPATPVANPVSQEQ---GQNPSSLSI 44
| : : : : : : : : : : : : : : : : : : : : : : :
DB 1335 LESLGYSLYASLGTAADFYTEHGKVTAVDWHFEAVDGECPQRSI 1380

RESULT 13
APAH_XANAC
ID APAH_XANAC STANDARD; PRT; 314 AA.
AC Q8PP27;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Bis(5'-nucleosyl)-tetraphosphatase, symmetrical (EC 3.6.1.41)
DE (Diadenosine tetraphosphatase) (AP4A hydrolase) (Diadenosine 5',5''-
DE p1,p4-tetraphosphate pyrophosphohydrolase).
GN APAH OR XAC0861.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Camarvan F., Cardozo J., Chamargo F., Clapina L.P.,
RA Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,

```

```

RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.";
RL Nature 417:459-463(2002).
CC -!- FUNCTION: Hydrolyzes diadenosine 5',5''-p1,p4-tetraphosphate to
CC yield ADP (By similarity).
CC -!- CATALYTIC ACTIVITY: P(1),P(4)-bis(5'-adenosyl)tetraphosphate +
CC H(2)O = 2 ADP.
CC -!- SIMILARITY: Belongs to the Ap4A hydrolase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE011718; AA035749.1; -
CC HAMAP; MF_00199; -; 1.
CC InterPro; IPR004617; APAH.
CC InterPro; IPR004843; M-peptidase.
CC InterPro; IPR006185; S/T phosphatase.
CC InterPro; IPR006186; T-kinase_APAH.
CC Pfam; PF00149; Metallophos; 1.
CC PRINTS; PR00114; STPHPTASE.
CC PRODOM; PD000252; T-kinase_APAH; 1.
CC TIGRFAMS; TIGR00668; APAH; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 314 AA; 34941 MW; 32981D55A639879B CRC64;

Query Match 21.4%; Score 58; DB 1; Length 314;
Best Local Similarity 32.6%; Pred. No. 8.2;
Matches 15; Conservative 5; Mismatches 26; Indels 0; Gaps 0;

QY 6 GFKLFATEATSDWLNANNVPATPVANPVSQEQNPSSLSIRKLRDG 51
| : : : : : : : : : : : : : : : : : : : : : : :
DB 250 GKKLTALQDLTDELRVVQVGNPTTHPKTAQRPPRRQRQGG 295

RESULT 14
PUS9_THEMA
ID PUS9_THEMA STANDARD; PRT; 452 AA.
AC Q9X0X6;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Bifunctional purine biosynthesis protein purH [includes:
DE Phosphoribosylamidoazolecarboxamide formyltransferase (EC 2.1.2.3)
DE (AICAR transformylase); IMP cyclohydrolase (EC 3.5.4.10) (Inosinase)
DE (IMP synthetase) (AIC)].
GN PURH OR TM1249.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Walek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.N., Sutton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Cotton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
CC -!- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + 5-amino-1-(5-

```

Db 1335 LESLGYSLYASLTADTYTGHGVKVTAVDWHFEAVDGECPQRSI 1380

RESULT 12

RT CAD. Controlled proteolysis, domains, and linkers.";

RL [8]

RN J. Biol. Chem. 267:7177-7184(1992).

RP 3D-STRUCTURE MODELING OF ATCase DOMAIN.

RX MEDLINE=91117244; PubMed=2006137;

RA Scully J.L., Evans D.R.;

RT "Comparative modeling of mammalian aspartate transcarbamylase.";

RL Proteins 9:191-206(1991).

CC -!- FUNCTION: THIS PROTEIN IS A "FUSION" PROTEIN ENCODING FOUR ENZYMIC ACTIVITIES OF THE PYRIMIDINE PATHWAY (GATase, CPase, ATCase AND DHase).

CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP + phosphate + L-glutamate + carbamoyl phosphate.

CC -!- CATALYTIC ACTIVITY: Carbamoyl phosphate + L-aspartate = phosphate + N-carbamoyl-L-aspartate.

CC -!- CATALYTIC ACTIVITY: (S)-dihydroorotate + H(2)O = N-carbamoyl-L-aspartate.

CC -!- COFACTOR: DHase ACTIVITY REQUIRES A ZINC ATOM.

CC -!- ENZYME REGULATION: ALLOSTERICALLY REGULATED AND CONTROLLED BY PHOSPHORYLATION. PRPP IS AN ACTIVATOR WHILE UMP IS AN INHIBITOR OF THE CPase REACTION.

CC -!- PATHWAY: FIRST TO THIRD STEP IN PYRIMIDINE BIOSYNTHESIS.

CC -!- SUBUNIT: Homohexamer.

CC -!- SUBCELLULAR LOCATION: Cytoplasmic.

CC -!- MISCELLANEOUS: GATase (GLUTAMINE AMIDOTRANSFERASE) AND CPase (CARBAMOYL PHOSPHATE SYNTHASE) FORM TOGETHER THE GLUTAMINE-DEPENDENT CPase (GD-CPase) (EC 5.3.5.5).

CC -!- SIMILARITY: THE CPase DOMAIN IS SIMILAR TO OTHER CPases.

CC -!- SIMILARITY: IN THE CENTRAL SECTION; BELONGS TO THE DHase FAMILY.

CC -!- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: J05503; AAA37062.1; -

DR EMBL: M28866; AAA37073.1; -

DR EMBL: M60078; AAA63617.1; -

DR EMBL: M11242; AAA37061.1; -

DR EMBL: M23652; AAA37064.1; -

DR EMBL: M21927; AAA37063.1; -

DR FIR: A38653; A23443.

DR HSSP: P00479; 3CSU.

DR InterPro: IPR006680; Amidohydro_1.

DR InterPro: IPR006220; Anth_synthII.

DR InterPro: IPR006130; Asp/Orn_Cotranf.

DR InterPro: IPR002082; Asp_carbmitransf.

DR InterPro: IPR006275; CarA_L_glu.

DR InterPro: IPR006274; CarA_small.

DR InterPro: IPR005483; CPase_L.

DR InterPro: IPR005479; CPase_L_D2.

DR InterPro: IPR005480; CPase_L_D3.

DR InterPro: IPR005481; CPase_L_N.

DR InterPro: IPR001317; CPase_GATase.

DR InterPro: IPR004474; CPase_sm_chain.

DR InterPro: IPR002193; Dihydroorotase.

DR InterPro: IPR000991; GATase_1.

DR InterPro: IPR004362; MGS_like.

DR InterPro: IPR006131; OTCace_O.

DR InterPro: IPR006132; OTCace_P.

DR InterPro: IPR004722; PyrC_multi.

DR InterPro: IPR005847; Urease.

DR Pfam: PF01979; Amidohydro_1; 1.

DR Pfam: PF00289; CPase_L_chain; 2.

DR Pfam: PF02786; CPase_L_D2; 2.

DR Pfam: PF02787; CPase_L_D3; 1.

DR Pfam: PF00988; CPase_sm_chain; 1.

DR Pfam: PF00117; GATase; 1.

DR Pfam: PF02142; MGS; 1.

DB 1335 LESLGYSLYASLTADTYTGHGVKVTAVDWHFEAVDGECPQRSI 1380

RT PRL_MESAU STANDARD; PRT; 2225 AA.

RX P08955; P70108;

DT 01-NOV-1988 (Rel. 09, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE CAD protein [Includes: Glutamine-dependent carbamoyl-phosphate synthase (EC 6.3.5.5); Aspartate carbamoyltransferase (EC 2.1.3.2); Dihydroorotase (EC 3.5.2.3)].

GN CAD.

OS Mesocricetus auratus (Golden hamster).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;

OC Mesocricetus.

OX NCBI_TaxID=10036;

RN [1]

RP SEQUENCE OF 1-169 FROM N.A.

RX MEDLINE=91139675; PubMed=1671675;

RA Bein K., Simmer J.P., Evans D.R.;

RT "Molecular cloning of a cDNA encoding the amino end of the mammalian multifunctional protein CAD and analysis of the 5'-flanking region of the CAD gene.";

RL J. Biol. Chem. 266:3791-3799(1991).

RN [2]

RP SEQUENCE OF 156-1455 FROM N.A.

RX MEDLINE=90285162; PubMed=1972379;

RA Simmer J.P., Kelly R.E., Rinker A.G. Jr., Scully J.L., Evans D.R.;

RT "Mammalian carbamyl phosphate synthetase (CPS). DNA sequence and evolution of the CPS domain of the Syrian hamster multifunctional protein CAD.";

RL J. Biol. Chem. 265:10395-10402(1990).

RN [3]

RP SEQUENCE OF 246-513 FROM N.A.

RX MEDLINE=88309082; PubMed=2900634;

RA Maley J.A., Davidson J.N.;

RT "Identification of the junction between the glutamine amidotransferase and carbamyl phosphate synthetase domains of the mammalian CAD protein.";

RL Biochem. Biophys. Res. Commun. 154:1047-1053(1988).

RN [4]

RP SEQUENCE OF 1391-1870 FROM N.A.

RX MEDLINE=90115834; PubMed=1967494;

RA Simmer J.P., Kelly R.E., Rinker A.G. Jr., Zimmermann B.H., Scully J.L., Kim H., Evans D.R.;

RT "Mammalian dihydroorotase: nucleotide sequence, peptide sequences, and evolution of the dihydroorotase domain of the multifunctional protein CAD.";

RL Proc. Natl. Acad. Sci. U.S.A. 87:174-178(1990).

RN [5]

RP SEQUENCE OF 1774-2225 FROM N.A.

RX MEDLINE=8928776; PubMed=2543974;

RA Simmer J.P., Kelly R.E., Scully J.L., Grayson D.R., Rinker A.G. Jr., Bergh S.T., Evans D.R.;

RT "Mammalian aspartate transcarbamylase (ATCase): sequence of the ATCase domain and interdomain linker in the CAD multifunctional polypeptide and properties of the isolated domain.";

RL Proc. Natl. Acad. Sci. U.S.A. 86:4382-4386(1989).

RN [6]

RP SEQUENCE OF 2074-2225 FROM N.A.

RX MEDLINE=85267690; PubMed=2862577;

RA Shigesada K., Stark G.R., Maley J.A., Niswander L.A., Davidson J.N.;

RT "Construction of a cDNA to the hamster CAD gene and its application toward defining the domain for aspartate transcarbamylase.";

RL Mol. Cell. Biol. 5:1735-1742(1985).

RN [7]

RP DOMAINS, AND PARTIAL SEQUENCE.

RX MEDLINE=92202288; PubMed=1348059;

RA Kim H., Kelly R.E., Evans D.R.;

RT "The structural organization of the hamster multifunctional protein

RESULT 10

PYRL_SCHPO STANDARD; PRF; 2244 AA.
 AC Q09794;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE URAL protein [Includes: Glutamine-dependent carbamoyl-phosphate
 synthase (EC 6.3.5.5); Aspartate carbamoyltransferase (EC 2.1.3.2)].
 GN URAL OR SPAC2267.06c.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 OX NCBI_TaxID=4896;
 RN [1]
 RN SEQUENCE OF 22-2244 FROM N.A.
 RP STRAIN=972;
 RC MEDLINE=96020160; PubMed=8590465;
 RA Lollmer M., Jaquet L., Nedeva T., Lacroute F., Potier S.,
 RA Souciet J.-L.;
 RA "As in Saccharomyces cerevisiae, aspartate transcarbamoylase is
 RT assembled on a multifunctional protein including a dihydroorotase-like
 RT cryptic domain in Schizosaccharomyces pombe.";
 RL Curr. Genet. 28:138-149 (1995).
 RN [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=972;
 RC MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouros J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin M., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders R., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volkart G., Aert R., Robben J., Grynolprez B.,
 RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Friz C., Holzer E., Mostl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Leinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakowski G.V., Ussery D., Barrell B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomyces pombe.";
 RL Nature 415:871-880 (2002).
 CC -!- FUNCTION: THIS PROTEIN IS A "FUSION" PROTEIN ENCODING THREE
 CC ENZYMAIC ACTIVITIES OF THE PYRIMIDINE PATHWAY (GATASE, CPSPASE,
 CC AND ATCPASE) (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
 CC phosphate + L-glutamate + carbamoyl phosphate.
 CC -!- CATALYTIC ACTIVITY: Carbamoyl phosphate + L-aspartate = phosphate
 CC + N-carbamoyl-L-aspartate.
 CC -!- PATHWAY: Pyrimidine biosynthesis; first step.
 CC -!- PATHWAY: Pyrimidine biosynthesis; second step.
 CC -!- MISCELLANEOUS: GATASE (GLUTAMINE AMIDOTRANSFERASE) AND CPSPASE
 CC (CARBAMOYL PHOSPHATE SYNTHASE) FORM TOGETHER THE
 CC GLUTAMINE-DEPENDENT CPSPASE (GD-CPSPASE) (EC 6.3.5.5).
 CC -!- MISCELLANEOUS: IN EUKARYOTES EC 6.3.5.5 IS SYNTHESIZED BY TWO
 CC PATHWAY-SPECIFIC (ARGININE AND PYRIMIDINE) UNDER SEPARATE CONTROL.
 CC -!- SIMILARITY: THE CPSPASE DOMAIN IS SIMILAR TO OTHER CPSPASES.
 CC -!- SIMILARITY: IN THE CENTRAL SECTION; BELONGS TO THE DHOASE FAMILY.
 CC DEFECTIVE DOMAIN.
 CC -!- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X81841; CAA57433.1; -;
 DR EMBL; Z54328; CAA91130.1; -;
 DR PIR; S65074; S65074.
 DR PIR; T11616; T11616.
 DR HSSP; P00968; 1A9X.
 DR GeneDB_Spombe; SPAC2267.06c; -;
 DR InterPro; IPR006220; Anth_synthII.
 DR InterPro; IPR006130; Asp/Ori_Cotranf.
 DR InterPro; IPR002082; Asp_carbMltransf.
 DR InterPro; IPR006275; CarA_L_glu.
 DR InterPro; IPR006274; CarA_small.
 DR InterPro; IPR005493; CPase_L.
 DR InterPro; IPR005479; CPase_L_D2.
 DR InterPro; IPR005480; CPase_L_D3.
 DR InterPro; IPR005481; CPase_L_N.
 DR InterPro; IPR001317; CPS_GATase.
 DR InterPro; IPR002474; CPSPase_sm_chain.
 DR InterPro; IPR000991; GATase_1.
 DR InterPro; IPR004362; MGS_like.
 DR InterPro; IPR006131; OTCace_O.
 DR InterPro; IPR006132; OTCace_P.
 DR Pfam; PF00289; CPSPase_L_chain; 2.
 DR Pfam; PF02786; CPSPase_L_D2; 2.
 DR Pfam; PF02787; CPSPase_L_D3; 1.
 DR Pfam; PF00986; CPSPase_sm_chain; 1.
 DR Pfam; PF00117; GATase; 1.
 DR Pfam; PF02142; MGS; 1.
 DR Pfam; PF00185; OTCace; 1.
 DR Pfam; PF02729; OTCace_N; 1.
 DR PRINTS; PR00097; ANTSYNTHASEII.
 DR PRINTS; PR00100; AOTCPASE.
 DR PRINTS; PR00098; CPSPASE.
 DR PRINTS; PR00099; CPSPGATASE.
 DR PRINTS; PR00096; GATASE.
 DR TIGSFams; TIGR00670; asp_carb_tr; 1.
 DR TIGSFams; TIGR01369; CPSPaseII_lrg; 1.
 DR TIGSFams; TIGR01368; CPSPaseII_small; 1.
 DR PROSITE; PS00097; CARBAMOYLTRANSFERASE; 1.
 DR PROSITE; PS00442; GATASE_TYPE_I; 1.
 DR PROSITE; PS00866; CPSPASE_1; 2.
 DR PROSITE; PS00867; CPSPASE_2; 2.
 KW Pyrimidine biosynthesis; Ligase; Transferase; Multifunctional enzyme.
 FT DOMAIN 1 437
 FT GATASE (GLUTAMINE AMIDOTRANSFERASE)
 FT (BY SIMILARITY).
 FT DOMAIN 438 477
 FT LINKER (BY SIMILARITY).
 FT DOMAIN 478 1514
 FT CPSPASE (CARBAMOYL-PHOSPHATE SYNTHASE)
 FT (BY SIMILARITY).
 FT DOMAIN 1515 1524
 FT LINKER (BY SIMILARITY).
 FT DOMAIN 1525 1853
 FT DEFECTIVE DHOASE DOMAIN
 FT (BY SIMILARITY).
 FT DOMAIN 1854 1935
 FT LINKER (BY SIMILARITY).
 FT DOMAIN 1936 2244
 FT ATCPASE (ASPARTATE TRANSCARBAMYLASE)
 FT (BY SIMILARITY).
 FT ACT_SITE 338 338
 FT GATASE (BY SIMILARITY).
 FT ACT_SITE 422 422
 FT GATASE (BY SIMILARITY).
 FT ACT_SITE 424 424
 FT GATASE (BY SIMILARITY).
 FT CONFLICT 336 338
 FT GIC -> RYF (IN REF. 1).
 FT CONFLICT 1035 1039
 FT CAVRA -> LOFAQ (IN REF. 1).
 FT CONFLICT 1409 1410
 FT EL -> DV (IN REF. 1).
 FT CONFLICT 1975 1975
 FT G -> E (IN REF. 1).
 FT CONFLICT 2002 2002
 FT G -> E (IN REF. 1).
 SQ SEQUENCE 2244 AA; 248306 MW; 5700D153B50CD3E9 CRC64;
 Query Match 23.2%; Score 63; DB 1; Length 2244;

RL Mol. Gen. Genet. 207:314-319(1987).
RN [3]
RP SEQUENCE OF 1268-2214 FROM N.A.
RX MEDLINE=89255278; PubMed=2498313;
RA Nagy M., le Gouar M., Potter S., Souciet J.-L., Herve G.;
RT "The primary structure of the aspartate transcarbamylase region of
RT the URA2 gene product in *Saccharomyces cerevisiae*. Features involved
RT in activity and nuclear localization.";
RL J. Biol. Chem. 264:8366-8374(1989).
RN [4]
RP SEQUENCE OF 1-276 FROM N.A.
RX STRAIN=S288C / FY1679;
RX MEDLINE=97103775; PubMed=8948101;
RA Cziepluch C., Kordes E., Pujol A., Janiaux J.-C.;
RT "Sequencing analysis of a 40.2 kb fragment of yeast chromosome X
RT reveals 19 open reading frames including URA2 (5' end), TRK1, PBS2,
RT SPI10, GCD14, RPE1, PHO86, NCA3, ASF1, COT7, GZF3, two trna genes,
RT three remnant delta elements and a Ty4 transposon.";
RL Yeast 12:1471-1474(1996).
RN [5]
RP SEQUENCE OF 175-2214 FROM N.A.
RX STRAIN=S288C / FY1679;
RX MEDLINE=96408771; PubMed=8813765;
RA Katsoulou C., Tzermia M., Tavernarakis N., Alexandraki D.;
RT "Sequence analysis of a 40.7 kb segment from the left arm of yeast
RT chromosome X reveals 14 known genes and 13 new open reading frames
RT including homologues of genes clustered on the right arm of
RT chromosome XI.";
RL Yeast 12:787-797(1996).
CC -!- FUNCTION: THIS PROTEIN IS A "FUSION" PROTEIN ENCODING THREE
CC ENZYMATIC ACTIVITIES OF THE PYRIMIDINE PATHWAY (GATASE, CPSASE,
CC AND ATCASE).
CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
CC phosphate + L-glutamate + carbamoyl phosphate
CC -!- CATALYTIC ACTIVITY: Carbamoyl phosphate + L-aspartate = phosphate
CC + N-carbamoyl-L-aspartate.
CC -!- PATHWAY: Pyrimidine biosynthesis; first step.
CC -!- PATHWAY: Pyrimidine biosynthesis; second step.
CC -!- MISCELLANEOUS: GATASE (GLUTAMINE AMIDOTRANSFERASE) AND CPSASE
CC (CARBAMOYL PHOSPHATE SYNTHASE) FORM TOGETHER THE
CC GLUTAMINE-DEPENDENT CPSASE (GD-CPSASE) (EC 6.3.5.5).
CC -!- MISCELLANEOUS: IN EUKARYOTES EC 6.3.5.5 IS SYNTHESIZED BY TWO
CC PATHWAY-SPECIFIC (ARGININE AND PYRIMIDINE) UNDER SEPARATE CONTROL.
CC -!- SIMILARITY: THE CPSASE DOMAIN IS SIMILAR TO OTHER CPSASES.
CC -!- SIMILARITY: IN THE CENTRAL SECTION; BELONGS TO THE DHOASE FAMILY.
CC -!- DEFECTIVE DOMAIN
CC -!- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M27174; AAA68280.1; ALT_INIT.
CC EMBL; X05553; CAA29068.1; -
CC EMBL; J04711; AAA35198.1; -
CC EMBL; Z49405; CAA89425.1; -
CC EMBL; X87371; CAA60825.1; -
CC EMBL; D28139; BAA05680.1; -
CC PIR; S56911; QZBYU2.
CC HSP; P00968; IA9X.
CC SGD; S000366; URA2.
CC GO; GO:0005737; C:cytoplasm; IDA.
CC GO; GO:0004070; F:aspartate carbamoyltransferase activity; IDA.
CC InterPro; IPR006220; Anth_synthil.
CC InterPro; IPR006130; Asp/Orn_Cotranf.
CC InterPro; IPR002082; Asp_carbamyltransf.
CC InterPro; IPR006275; CarA_L_glu.
CC InterPro; IPR006274; CarA_small.
CC InterPro; IPR005483; CPase_L.

DR InterPro; IPR005479; CPase_L_D2.
DR InterPro; IPR005480; CPase_L_D3.
DR InterPro; IPR005481; CPase_L_N.
DR InterPro; IPR001317; CPS_GATase.
DR InterPro; IPR002474; CPSase_sm_chain.
DR InterPro; IPR000991; GATase_1.
DR InterPro; IPR004362; MGS_like.
DR InterPro; IPR006131; OTCace_O.
DR InterPro; IPR006132; OTCace_P.
DR Pfam; PF00289; CPSase_L_chain; 2.
DR Pfam; PF02786; CPSase_L_D2; 2.
DR Pfam; PF02787; CPSase_L_D3; 1.
DR Pfam; PF00988; CPSase_sm_chain; 1.
DR Pfam; PF00117; GATase; 1.
DR Pfam; PF02142; MGS; 1.
DR Pfam; PF00185; OTCace; 1.
DR Pfam; PF02729; OTCace_N; 1.
DR PRINTS; PR00097; ANTSNHASEII.
DR PRINTS; PR00100; AOTCASE.
DR PRINTS; PR00098; CPSASE.
DR PRINTS; PR00099; CPSGATASE.
DR PRINTS; PR00096; GATASE.
DR TIGRFAMS; TIGR00670; asp_carb_tr; 1.
DR TIGRFAMS; TIGR01369; CPSaseII_ltr; 1.
DR TIGRFAMS; TIGR01368; CPSaseII_small; 1.
DR PROSITE; PS00097; CARBAMOYLTRANSFERASE; 1.
DR PROSITE; PS00442; GATASE_TYPE_I; 1.
DR PROSITE; PS00866; CPSASE_1; 2.
DR PROSITE; PS00867; CPSASE_2; 2.
KW Pyrimidine biosynthesis; Ligase; Transferase; Multifunctional enzyme.
FT DOMAIN 1 400
FT DOMAIN 401 440
FT DOMAIN 441 1482
FT DOMAIN 1483 1492
FT DOMAIN 1493 1821
FT DOMAIN 1822 1909
FT DOMAIN 1910 2214
FT ACT_SITE 302 302
FT ACT_SITE 386 386
FT ACT_SITE 388 388
FT CONFLICT 86 86
FT CONFLICT 123 123
FT CONFLICT 123 123
FT CONFLICT 250 257
FT CONFLICT 270 270
FT CONFLICT 313 314
FT CONFLICT 372 373
FT CONFLICT 394 402
FT CONFLICT 431 433
FT CONFLICT 482 482
FT CONFLICT 485 485
FT CONFLICT 492 492
FT CONFLICT 501 510
FT CONFLICT 852 852
FT CONFLICT 1411 1412
FT CONFLICT 1582 1582
FT CONFLICT 1588 1588
FT CONFLICT 1592 1592
FT CONFLICT 1595 1595
FT CONFLICT 1937 1937
FT CONFLICT 1997 1997
FT CONFLICT 2039 2039
FT CONFLICT 2158 2165
SQ SEQUENCE 2214 AA; 245124 MW; 4CA58304DAECAD21 CRC64;
GATASE (GLUTAMINE AMIDOTRANSFERASE).
LINKER.
CPSASE (CARBAMOYL-PHOSPHATE SYNTHASE).
LINKER.
DEFECTIVE DHOASE DOMAIN.
ATCASE (ASPARTATE TRANSCARBAMYLASE).
GATASE (BY SIMILARITY).
GATASE (BY SIMILARITY).
GATASE (BY SIMILARITY).
H -> D (IN REF. 2).
R -> A (IN REF. 1 AND 2).
ELKVPWN -> RIESCMD (IN REF. 1 AND 2).
I -> Y (IN REF. 1 AND 2).
GA -> VO (IN REF. 1 AND 2).
GI -> RE (IN REF. 1 AND 2).
RQTEFLDY -> EIQNSCLT (IN REF. 1 AND 2).
KAH -> QGT (IN REF. 1 AND 2).
I -> T (IN REF. 2).
I -> N (IN REF. 2).
A -> G (IN REF. 2).
D -> V (IN REF. 1).
EV -> S (IN REF. 1 AND 3).
I -> M (IN REF. 1 AND 3).
N -> K (IN REF. 1 AND 3).
V -> G (IN REF. 1 AND 3).
S -> A (IN REF. 1 AND 3).
A -> R (IN REF. 3).
T -> I (IN REF. 3).
H -> L (IN REF. 3).
KILAHKE -> VRSWHTQCK (IN REF. 3).
24.0%; Score 65; DB 1; Length 2214;
Best Local Similarity 45.8%; Pred. No. 10;
Matches 11; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

QY 1 QLNHEGFKLFATEATSDWLNANNV 24

Db 1382 KLYNGYKLFATSGTADFLSEHGI 1405

CARB_BACSU STANDARD; PRT: 1071 AA.

P25994;
01-MAY-1992 (Rel. 22, Created)
01-MAY-1992 (Rel. 22, Last sequence update)
28-FEB-2003 (Rel. 41, Last annotation update)
Carbamoyl-phosphate synthase, pyrimidine-specific, large chain
(RC 6.3.5.5) (Carbamoyl-phosphate synthetase ammonia chain).
PYRAB.
Bacillus subtilis.
Bacteria: Firmicutes; Bacillales; Bacillaceae; Bacillus.
NCBI_TaxID=1423;
[1]
SEQUENCE FROM N.A.
MEDLINE=91225016; PubMed=1709162;
Quinn C.L., Stephenson B.T., Switzer R.L.;
"Functional organization and nucleotide sequence of the Bacillus
subtilis pyrimidine biosynthetic operon.";
J. Biol. Chem. 266:9113-9127(1991).
[2]
SEQUENCE FROM N.A.
STRAIN=168;
MEDLINE=98044033; Pubmed=9384377;
Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borcher S.,
Borriss K., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
Choi S.K., Codani J.J., Connerion I.F., Cummings N.J., Daniel R.A.,
Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmeron P.T.,
Entian K.D., Errington J., Fabret C., Ferrari E., Poulger D.,
Fritz C., Fujita M., Fujita Y., Funa S., Galizzi A., Galleron N.,
Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
Hilbert H., Hollappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
Joris B., Kasamata D., Kasahara Y., Klaerr-Blauchard M., Klein C.,
Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
Kurita K., Lapidus A., Lardinios S., Lauber J., Lazarevic V.,
Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
Presecan E., Puigc J., Purnelle B., Rapoport G., Rey M., Reynolds S.,
Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadalie Y.,
Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
Sorokin A., Taconi E., Takagi T., Takahashi H., Takemaru K.,
Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
Tosato V., Uchiyama S., Vandebol M., Vannier F., Vassarotti A.,
Viari A., Wambutt R., Wedler E., Wedler H., Weitzengraber T.,
Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
"The complete genome sequence of the Gram-positive bacterium Bacillus
subtilis";
Nature 390:249-256(1997).
CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
phosphate + L-glutamate + carbamoyl phosphate.
CC -!- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -!- PATHWAY: Pyrimidine biosynthesis; first step.
CC -!- SUBUNIT: Composed of two chains; the small (or glutamine) chain
promotes the hydrolysis of glutamine to ammonia, which is used by
the large (or ammonia) chain to synthesize carbamoyl phosphate.
CC -!- SIMILARITY: BELONGS TO THE CARB FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).

EMBL: M59757; AAA21270.1; -
DR EMBL: Z59112; CAB13426.1; -
DR FIR: F39845; F39845.


```

FT REPEAT      416
FT REPEAT     967
FT ACT_SITE   290
FT ACT_SITE   290
SQ SEQUENCE    1496 AA; 163767 MW; 2360BF05D21B059E CRC64;

Query Match      68.1%; Score 184.5; DB 1; Length 1496;
Best Local Similarity 72.0%; Pred. No. 1.3e-14;
Matches 36; Conservative 8; Mismatches 5; Indels 1; Gaps 1;

QY 2 LHNEGFKLFAATSDWLNANNVPATPVAMPVWPQEQGNFSLSSIRKLIRDG 51
DB 1378 LKDEGFKLVAETADWLNANDITAPVAMPVWPQEQGS-GPSSIYKLIGES 1426
      :|:|||||||:|||||||: |||||: | ||| |:|
      :|:|||||||:|||||||: |||||: | ||| |:|

RESULT 4
CAB2_METJA STANDARD; PRT; 618 AA.
ID CAB2_METJA
AC Q58776;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain, C-terminal section
DE (EC 6.3.5.5) (Carbamoyl-phosphate synthetase ammonia chain).
DE CARB2 OR MJ1381.
GN Methanococcus jannaschii.
OC Archaea: Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A. 2661 / AFCC 43067;
RC STRAIN=JAL-1 / DSM 2661 / AFCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.P., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.D., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.K., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenn H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RA jannaschii".
RL Science 273:1058-1073(1996).
CC -I- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
CC phosphate + L-glutamate + carbamoyl phosphate.
CC -I- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -I- PATHWAY: Arginine biosynthesis.
CC -I- PATHWAY: Pyrimidine biosynthesis; first step.
CC -I- SUBUNIT: Composed of two chains; the small (or glutamine) chain
CC promotes the hydrolysis of glutamine to ammonia, which is used by
CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By
CC similarity).
CC -I- SIMILARITY: BELONGS TO THE CARB FAMILY. C-TERMINAL SECTION.
CC -I- CAUTION: Sequence of carB is split into two genes in M.jannaschii
CC (MJ1378 and MJ1381).
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation at
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; U67578; AB99391.1; -.
DR F1R; D64472; D64472.
DR HSP; P00968; LBXR.
DR TIGR; MJ1381; -.
DR HAMAP; MF_01210; atypical; 1.
DR InterPro; IPR005483; CPhase_L.
DR InterPro; IPR005479; CPhase_LD2.
DR InterPro; IPR005480; CPhase_L_D3.
DR InterPro; IPR005481; CPhase_L_N.

```


CC symptoms are vomiting in infancy, protein intolerance,
 CC intermittent ataxia, seizures, lethargy, and mental retardation.
 CC -!- SIMILARITY: TO OTHER CARBAMOYL-PHOSPHATE SYNTHETASES, ALSO
 CC CONTAINS A GLUTAMINE AMIDOTRANSFERASE-LIKE DOMAIN BUT THE CYSTEINE
 CC RESIDUE WHICH IS ESSENTIAL FOR AMIDO TRANSFERASE ACTIVITY, IS
 CC MISSING IN THIS PROTEIN SO THIS DOMAIN IS NO LONGER ACTIVE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D90282; BAA14328.1; -
 CC EMBL: Y15793; CAA75785.1; -
 CC EMBL: AF154830; AAD38072.1; -
 CC PIR: JQ1348; JQ1348.
 CC HSRP: P00968; IAXX.
 CC Genew: HGNC:2323; CPS1.
 CC GK: P31327; -
 CC MIM: 237300; -
 CC GO: GO:0005739; C:mitochondrion; TAS.
 CC GO: GO:0004087; F:carbamoyl-phosphate synthase (ammonia) acti. .; TAS.
 CC InterPro: IPR006275; CarA_L.glu.
 CC InterPro: IPR006274; CarA_small.
 CC InterPro: IPR005483; CPase_L.
 CC InterPro: IPR005479; CPase_L.D2.
 CC InterPro: IPR005480; CPase_L.D3.
 CC InterPro: IPR005481; CPase_L.N.
 CC InterPro: IPR001317; CPS_Gatase.
 CC InterPro: IPR002474; CPase_sm_chain.
 CC InterPro: IPR000991; Gatase_1.
 CC InterPro: IPR004362; MGS_like.
 CC Pfam: PF00289; CPase_L_chain; 2.
 CC Pfam: PF02786; CPase_L_D2; 2.
 CC Pfam: PF02787; CPase_L_D3; 1.
 CC Pfam: PF00988; CPase_sm_chain; 1.
 CC Pfam: PF00117; Gatase; 1.
 CC Pfam: PF02142; MGS; 1.
 CC PRINTS: PR00098; CPSASE.
 CC PRINTS: PR00099; CPSGATASE.
 CC PRINTS: PR00096; GATASE.
 CC TIGRFAMS: TIGR01369; CPaseII_lrg; 1.
 CC TIGRFAMS: TIGR01368; CPaseII_small; 1.
 CC PROSITE: PS00866; CPSASE_1; 2.
 CC PROSITE: PS00867; CPSASE_2; 2.
 KW Ligase; Allosteric enzyme; Repeat; Transit peptide; Mitochondrion;
 KW ATP-binding; Urea cycle; Polymorphism; Disease mutation.
 FT TRANSIT 1 38
 FT CHAIN 39 1500
 FT NP_BIND 571 626
 FT NP_BIND 718 768
 FT NP_BIND 1113 1171
 FT NP_BIND 1259 1302
 FT DOMAIN 39 219
 FT DOMAIN 220 410
 FT DOMAIN 411 1500
 FT REPEAT 419 876
 FT REPEAT 970 1410
 FT VARIANT 337 337
 FT VARIANT 344 344
 FT VARIANT 544 544
 FT VARIANT 111 111
 FT CONFLICT 279 279
 FT CONFLICT 338 338
 FT CONFLICT 718 722

FT CONFLICT 729 729 A -> T (IN REF. 1).
 FT CONFLICT 749 749 E -> G (IN REF. 1).
 FT CONFLICT 1161 1162 EH -> AT (IN REF. 1).
 FT CONFLICT 1204 1205 GD -> EN (IN REF. 1).
 FT CONFLICT 1254 1254 I -> N (IN REF. 1).
 FT CONFLICT 1266 1266 F -> S (IN REF. 1).
 FT CONFLICT 1283 1283 M -> L (IN REF. 1).
 FT CONFLICT 1303 1303 A -> V (IN REF. 1).
 FT CONFLICT 1406 1406 T -> N (IN REF. 1).
 SQ SEQUENCE 1500 AA; 164939 MW; E53A22D77563961D CRC64;
 Query Match 100.08; Score 271; DB 1; Length 1500;
 Best local similarity 100.08; Pred. No. 2.8e-25;
 Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 QLNHEGFKLFATEATSDWLNANNVPATPVANPQEQNPSSLSIRKLIRG 51
 DB 1380 QLNHEGFKLFATEATSDWLNANNVPATPVANPQEQNPSSLSIRKLIRG 1430
 RESULT 2
 CPSM_RAT STANDARD; PRT; 1500 AA.
 ID AC P07756;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Carbamoyl-phosphate synthase (ammonia), mitochondrial precursor
 DE (EC 6.3.4.16) (Carbamoyl-phosphate synthetase I) (CPSASE I).
 GN CPS1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85261323; PubMed=2991241;
 RA Nyunova H., Broglie K.E., Widgren E.E., Lusty C.J.;
 RT "Characterization and derivation of the gene coding for mitochondrial
 carbamyl phosphate synthetase I of rat."
 RL J. Biol. Chem. 262:10415-10418(1987).
 RN [2]
 RP SEQUENCE OF 1-42 FROM N.A.
 RX MEDLINE=87280088; PubMed=3038878;
 RA Ladace M., Howell B.W., Burak R., Lusty C.J., Shore G.C.;
 RT "Rat carbamyl-phosphate synthetase I gene. Promoter sequence and
 tissue-specific transcriptional regulation in vitro."
 RL J. Biol. Chem. 262:10415-10418(1987).
 CC -!- FUNCTION: INVOLVED IN THE UREA CYCLE OF UREOTELIC ANIMALS WHERE
 THE ENZYME PLAYS AN IMPORTANT ROLE IN REMOVING EXCESS AMMONIA
 FROM THE CELL.
 CC -!- CATALYTIC ACTIVITY: 2 ATP + NH(3) + CO(2) + H(2)O = 2 ADP +
 phosphate + carbamoyl phosphate.
 CC -!- ENZYME REGULATION: REQUIRES N-ACETYLGLUTAMATE AS AN ALLOSTERIC
 ACTIVATOR.
 CC -!- SUBCELLULAR LOCATION: Mitochondrial.
 CC -!- TISSUE SPECIFICITY: PRIMARILY IN THE LIVER AND SMALL INTESTINE.
 CC -!- PTM: 50% OF THE MATURE PROTEIN THAT WAS ISOLATED HAD LEU 39 AS ITS
 N-TERMINAL RESIDUE AND 50% HAD SER 40 SUGGESTING TWO ADJACENT
 PROCESSING SITES. HOWEVER, THE POSSIBILITY OF PROTEOLYTIC REMOVAL
 OF LEU 39 DURING THE ISOLATION OF THE ENZYME CANNOT BE EXCLUDED.
 CC -!- SIMILARITY: TO OTHER CARBAMOYL-PHOSPHATE SYNTHETASES. ALSO
 CONTAINS A GLUTAMINE AMIDOTRANSFERASE-LIKE DOMAIN BUT THE CYSTEINE
 RESIDUE WHICH IS ESSENTIAL FOR AMIDO TRANSFERASE ACTIVITY, IS
 MISSING IN THIS PROTEIN SO THIS DOMAIN IS NO LONGER ACTIVE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2003, 02:03:24 : Search time 14 Seconds
(without alignments)
171.312 Million cell updates/sec

Title: US-09-585-077C-4_COPY_1380_1430

Perfect score: 271

Sequence: 1 QHNEGKLFATATSDWLN.....PSQEGQNPSLSIRKLIRDG 51

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	271	100.0	1500	1	CPSM_HUMAN
2	271	100.0	1500	1	CPSM_RAT
3	184.5	68.1	1496	1	CPSM_RANCA
4	82	30.3	618	1	CAB2_METJA
5	74.5	27.5	2185	1	PRL1_DICDI
6	69.5	25.6	1115	1	CARB_MYCTU
7	69	25.5	1102	1	CARB_STRCO
8	68.5	25.3	1071	1	CARB_BAGSU
9	65	24.0	2214	1	PYR1_YEAST
10	63	23.2	2244	1	PYR1_SCHPO
11	59.5	22.0	2225	1	PYR1_HUMAN
12	59.5	22.0	2225	1	PYR1_MESAU
13	58	21.4	314	1	APAH_XANAC
14	58	21.4	452	1	PUR9_THEMA
15	58	21.4	591	1	ALU8_HUMAN
16	57.5	21.2	264	1	CRK1_SALTY
17	57.5	21.2	457	1	BAG4_MOUSE
18	57	21.0	2242	1	PYR1_SQUAC
19	56.5	20.8	811	1	RFX1_YEAST
20	56.5	20.8	1121	1	CARB_MYCLE
21	56	20.7	1162	1	CARB_BRUME
22	55	20.3	209	1	LOBL_RALSO
23	55	20.3	254	1	VSPA_SOYBN
24	55	20.3	291	1	S2SK_SOYBN
25	55	20.3	441	1	PAC1_MOUSE
26	55	20.3	441	1	PAC1_RAT
27	55	20.3	444	1	PAC1_HUMAN
28	55	20.3	521	1	YF92_MYCPN
29	55	20.3	1113	1	CARB_CORGL
30	54.5	20.1	449	1	VPT1_CAV82
31	54.5	20.1	514	1	GSH1_BRAJU
32	54.5	20.1	662	1	YFL8_YEAST
33	54.5	20.1	1062	1	CARB_BACHD

RESULT 1

ID	CPSM_HUMAN	STANDARD;	PRT;	1500 AA.
AC	P31327; O43774;			
DT	01-JUL-1993 (Rel. 26, Created)			
DT	15-DEC-1998 (Rel. 37, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Carbamoyl-phosphate synthase [ammonia], mitochondrial precursor (EC 6.3.4.16) (Carbamoyl-phosphate synthetase I) (CPSase I).			
GN	CPSI.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Liver;			
RX	MEDLINE=92084128; PubMed=1840546;			
RA	Haraguchi Y., Uchino T., Takiguchi M., Endo F., Mori M.,			
RA	Matsuda I.;			
RT	"Cloning and sequence of a cDNA encoding human carbamyl phosphate synthetase I: molecular analysis of hyperammonemia.";			
RT	Hum. Mutat. 12:206-211(1998).			
RL	Gene 107:335-340(1991).			
RN	[2]			
RP	SEQUENCE FROM N.A., VARIANT CPSI DEFICIENCY MET-544, AND VARIANT			
RC	ALA-344.			
RC	TISSUE=Liver;			
RX	MEDLINE=98375696; PubMed=9711878;			
RA	Finckh U., Kohlschuetter A., Schaefer H., Sperhake K., Colombo J.-P.,			
RA	Gal A.;			
RT	"Prenatal diagnosis of carbamoyl phosphate synthetase I deficiency by identification of a missense mutation in CPSI.";			
RT	Hum. Mutat. 12:206-211(1998).			
RL	[3]			
RN	SEQUENCE FROM N.A.			
RP	Summar M.;			
RA	Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.			
RN	[4]			
RP	VARIANT CPSI DEFICIENCY ARG-337.			
RX	MEDLINE=21367742; PubMed=1147210;			
RA	Aoshima T., Kajita M., Sekido Y., Kikuchi S., Yasuda I., Saheki T.,			
RA	Watanabe K., Shimokata K., Niwa T.;			
RT	"Novel mutations (H337R and 238-362del) in the CPSI gene cause carbamoyl phosphate synthetase I deficiency.";			
RT	Hum. Hered. 52:99-101(2001).			
RL	Hum. Hered. 52:99-101(2001).			
CC	-!- FUNCTION: INVOLVED IN THE UREA CYCLE OF UREOTELIC ANIMALS WHERE			
CC	THE ENZYME PLAYS AN IMPORTANT ROLE IN REMOVING EXCESS AMMONIA			
CC	FROM THE CELL.			
CC	-!- CATALYTIC ACTIVITY: 2 ATP + NH(3) + CO(2) + H(2)O = 2 ADP +			
CC	phosphate + carbamoyl phosphate.			
CC	-!- ENZYME REGULATION: REQUIRES N-ACETYLGLUTAMATE AS AN ALLOSTERIC			
CC	ACTIVATOR.			
CC	-!- SUBCELLULAR LOCATION: Mitochondrial.			
CC	-!- TISSUE SPECIFICITY: PRIMARILY IN THE LIVER AND SMALL INTESTINE.			
CC	-!- DISEASE: Defects in CPSI are the cause of an autosomal recessive			
CC	metabolic disorder that cause a type of hyperammonemia. Clinical			

34	54.5	20.1	1228	1	SLAP_BACST	P35825 bacillus st
35	54	19.9	423	1	EPD1_RALSO	P58991 ralstonia s
36	54	19.9	488	1	PAC2_RAT	O9q17 rattus norv
37	54	19.9	1085	1	CUT7_SCHPO	P24339 schizosacch
38	54	19.9	2280	1	COAC_SCHPO	P78820 schizosacch
39	53.5	19.7	199	1	PEXH_YEAST	P40155 saccharomyc
40	53.5	19.7	270	1	TCMP_STRGA	P39887 streptomyce
41	53.5	19.7	323	1	VALL_PASVK	Q00338 panicum str
42	53.5	19.7	462	1	MYCN_MOUSE	P03966 mus musculu
43	53.5	19.7	1092	1	CARB_ZYMMO	O50236 zymomonas m
44	53	19.6	288	1	MRGX_HUMAN	Q15014 homo sapien
45	53	19.6	772	1	LMBT_HUMAN	Q9y468 homo sapien

ALIGNMENTS

A:Reference number: S62445
 A:Accession: S62450
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 'MSGLLPSLSSSPPLYQSEALG', 1-314, 'GIC', 318-1013, 'CAVR', 1017, 1019-1387, 'EL', 1390
 A:Cross-references: EMBL:Z54328; NID:g1009451; PIDN:CRA91130.1; PID:g1009456
 C:Genetics:
 A:Gene: URAL
 A:Map position: 1L
 C:Superfamily: rudimentary enzyme; aspartate/ornithine carbamoyltransferase homology; Ba
 carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain homology; carbamoyl-phos
 C:Keywords: ligase; multifunctional enzyme; pyrimidine nucleotide biosynthesis; transfer
 F:40-1498/Domain: carbamoyl-phosphate synthase (ammonia) homology <CPA>
 F:40-1498/Domain: carbamoyl-phosphate synthase (glutamine-hydrolyzing) small chain homolo
 F:244-419/Domain: trpG homology <TRG>
 F:244-419/Domain: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain homoc
 F:455-1495/Domain: biotin carboxylase homology <BC1>
 F:455-1437/Domain: biotin carboxylase homology <BC2>
 F:994-1437/Domain: biotin carboxylase homology <BC2>
 F:1511-1836/Domain: Bacillus dihydroorotase homology <DHO>
 F:1916-2216/Domain: aspartate/ornithine carbamoyltransferase homology <ACT>

Query Match 23.2%; Score 63; DB 2; Length 2223;
 Best Local Similarity 35.6%; Pred. No. 32;
 Matches 16; Conservative 10; Mismatches 17; Indels 2; Gaps 2;

QY 1 QLNHGFKLFATETSDWLNANNVPATPVA-WPQEGONP-SLSS 43
 DB 1395 KLYENNYIFATAGTSDYFMESGVPCKYLADLPFAEANNYSLSA 1439

Search completed: September 13, 2003, 02:16:43
 Job time : 24.5 secs

A;Cross-references: GB:AE008688; PIDN:AAL42195.1; PID:g17739586; GSPDB:GN00186
A;Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atul183
A:Map position: circular chromosome
C:Superfamily: phage T7 DNA-directed RNA polymerase

Query Match 23.6%; Score 64; DB 2; Length 863;
Best Local Similarity 30.2%; Pred.No. 7.6;
Matches 19; Conservative 7; Mismatches 23; Indels 14; Gaps 2;

QY 3 HNEGKFLFATATSDWL---NANNVPATPVAMPSEOGQNPSLSIR-----KLI 48
|| : ||: | ||| | : | || | : | | : | : ||
Db 221 HNNTIYVVAKETLDWLATNSRLAPSLPYLTLPVPRPWTSPFRGGYWSGRVNRLI 280
|| : ||: | ||| | : | || | : | | : | : ||

QY 49 RDG 51
: |
Db 281 KTG 283

RESULT 14
B97504
RNA polymerase [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C:Species: Agrobacterium tumefaciens
C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
C:Accession: B97504
R:Goodner, B.; Hinkley, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz,
Science 294, 2323-2328, 2001
A:title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A:Reference number: A97359; MID:21608551; PMID:11743194
A:Accession: B97504
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-863 <XUR>
A:Cross-references: GB:AE007869; PIDN:AK86987.1; PID:g15156227; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C_2186
A:Map position: circular chromosome
C:Superfamily: phage T7 DNA-directed RNA polymerase

Query Match 23.6%; Score 64; DB 2; Length 863;
Best Local Similarity 30.2%; Pred.No. 7.6;
Matches 19; Conservative 7; Mismatches 23; Indels 14; Gaps 2;

QY 3 HNEGKFLFATATSDWL---NANNVPATPVAMPSEOGQNPSLSIR-----KLI 48
|| : ||: | ||| | : | || | : | | : | : ||
Db 221 HNNTIYVVAKETLDWLATNSRLAPSLPYLTLPVPRPWTSPFRGGYWSGRVNRLI 280
|| : ||: | ||| | : | || | : | | : | : ||

QY 49 RDG 51
: |
Db 281 KTG 283

RESULT 15
S65074
pyrimidine synthesis protein URAL - fission yeast (Schizosaccharomyces pombe)
N:Alternate names: protein SPAC22G7.06c
N:Contains: aspartate carbamoyltransferase (EC 2.1.3.2); carbamoyl-phosphate synthase
C:Species: Schizosaccharomyces pombe
C>Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 20-Jun-2000
C:Accession: S65074; S62450
R:Löhlmer, M.; Jaquet, I.; Nedeva, T.; Lacroute, F.; Potier, S.; Souciot, J.L.
Curr. Genet. 28, 138-149, 1995
A:title: As in Saccharomyces cerevisiae, aspartate transcarbamoylase is assembled on
A:Reference number: S65074; MID:196020160; PMID:8590465
A:Accession: S65074
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2223 <LOI>
A:Cross-references: EMBL:X01841; NID:g1154951; PIDN:CAAS57433.1; PID:g1154952
R:Badcock, K.; Churcher, C.M.
submitted to the EMBL Data Library, October 1995

Db ::::||||| | |||| | : :::::|||||:
61 RVHSEGFVFLQNAQDWLAFGRSHDMLPIAMPTR 95

RESULT 10
A70990
carbamoyl-phosphate synthase (glutamine-hydrolyzing) (EC 6.3.5.5) large chain [similarit
C:Species: Mycobacterium tuberculosis
C:date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 08-Sep-2000
C:Accession: A70990
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
C:Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Soares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; PMID:98295987; PMID:9634230
A:Accession: A70990
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1115 <COL>
A:Cross-references: GB:I281011; GB:AL123456; NID:g3242274
A:Experimental source: strain H37Rv
A>Note: This ORF is not annotated in GenBank entry M7C2B12, release 106
C:Genetics:
A:Gene: Rv1384; carB
A:Start codon: GTG
C:Superfamily: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain; biotin
C:Keywords: ligase
F:9-1089/Domain: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain homolo

Query Match 25.6%; Score 69.5; DB 2; Length 1115;
Best Local Similarity 35.2%; Pred. No. 2.1;
Matches 19; Conservative 13; Mismatches 17; Indels 5; Gaps 3;

QY 1 OLHNHGKFLFAETSDMLNANNVPAPVA--WPSEGQGNPSLSSTKLIRDG 51
992 RLADLGHFVLATGTAEMLRNGICDDVKHFPAQPCR-PTMSAV-DAIRAG 1043

Db
RESULT 11
F39845
carbamoyl-phosphate synthase (glutamine-hydrolyzing) (EC 6.3.5.5), pyrimidine-repressib
N:Alternate names: carbamoyl-phosphate synthetase (catalytic chain) pyrAB
C:Species: Bacillus subtilis
C:date: 20-Mar-1992 #sequence_revision 20-Mar-1992 #text_change 02-Sep-2000
C:Accession: F39845; A69686
R:Quinn, C.L.; Stephenson, B.T.; Switzer, R.L.
J. Biol. Chem. 266, 9113-9127, 1991
A:title: Functional organization and nucleotide sequence of the Bacillus subtilis pyrim
A:Reference number: A39845; PMID:91225016; PMID:1709162
A:Accession: F39845
A:Molecule type: DNA
A:Residues: 1-1071 <QUI>
A:Cross-references: GB:M59757; NID:g4887706; PIDN:AAA21270.1; PID:g143390
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.: Bron, S.; Branellet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chid
A.: Ehrlich, S.D.; Emerson, P.T.; Enian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Folger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallen
leach, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Hollappel, S.; Hosono, S.; Huilo, M.F.
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauele
Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetell
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadade, I.; Sato, T.; Scanlon
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron
akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terstra, P.; Togonoi, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Zanchin, A.
A:title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; PMID:198044033; PMID:9384377
A:Accession: A69686
A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA
A:Residues: 1-1071 <UN>
A:Cross-references: GB:Z99112; GB:AL009126; NID:g2633902; PIDN:CAE13426.1; PID:g263393
A:Experimental source: strain t68
C:Genetics:
A:Gene: pyrAB
C:Function:
A:Pathway: glutamate metabolism; pyrimidine nucleotide biosynthesis
C:Superfamily: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain; biot
C:Keywords: ligase; pyrimidine nucleotide biosynthesis
F:8-1047/Domain: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain hom
F:9-467/Domain: biotin carboxylase homolog <BC1>
F:553-985/Domain: biotin carboxylase homology <EC2>

Query Match 25.3%; Score 68.5; DB 2; Length 1071;
Best Local Similarity 31.4%; Pred No. 2.7;
Matches 16; Conservative 11; Mismatches 19; Indels 5; Gaps 1;

OY 1 QLNHEGFKLPATEQSDWLNANNVPATPVAMPQSQGPNSSIRKLIRDG 51
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 956 RFHAGYNILATEGTAGLYKEASIPAKVVGKIGDGNP-----LLDVIRNG 1001

RESULT 12
QZBRU2
pyrimidine synthesis protein URA2 - yeast (Saccharomycetes cerevisiae)
N:Alternate names: protein J0686; protein YJL130C
N:Contains: aspartate carbamoyltransferase (EC 2.1.1.3.2); carbamoyl-phosphate synthase
C:Species: Saccharomycetes cerevisiae
C:Date: 30-Sep-1991 #sequence revision 08-Sep-1995 #text change 12-Nov-1999
C:Accession: S56911; S56912; S03767; S03859; S55182; S13358; S71667; S22790; A33820;
R:Ziepluch, C.; Kordes, E.; Pujol, A.; Jauniaux, J.C.
A:Submitted to the Protein Sequence Database, September 1995
A:Reference number: S56891
A:Accession: S56911
A:Molecule type: DNA
A:Residues: 1-276 <X1>
A:Cross-references: EMBL:Z49405; GSPDB:GN00010; MIPS:YJL130C
R:Katsoulou, C.; Tzermia, M.; Alexandraki, D.
Submitted to the Protein Sequence Database, September 1995
A:Reference number: S56912
A:Accession: S56912
A:Molecule type: DNA
A:Residues: 175-2214 <KAT>
A:Cross-references: EMBL:Z49405; GSPDB:GN00010; MIPS:YJL130C
R:Souciet, J.L.; Nagy, M.; le Gouar, M.; Lacroute, F.; Potier, S.
Gene 79, 59-70, 1989
A:Title: Organization of the yeast URA2 gene: identification of a defective dihydroor
A:Reference number: S05766; MUID:89378778; PMID:2570735
A:Accession: S05767
A:Molecule type: DNA
A:Residues: 1-122, 'A', 124-249, 'RI', 250, 'SCSMD', 258-269, 'Y', 271-312, 'VQ', 315-371, 'RF', 'A'
A': 1596-2214 <SOUI>
A:Cross-references: EMBL:M27174
R:Souciet, J.L.; Potier, S.; Hubert, J.C.; Lacroute, F.
Mol. Genet. 207, 314-319, 1987
A:Title: Nucleotide sequence of the pyrimidine specific carbamoyl phosphate synthetas
A:Reference number: S03859; MUID:87286375; PMID:3038294
A:Accession: S03859
A:Molecule type: DNA
A:Residues: 1-85, 'D', 87-122, 'A', 124-249, 'RI', 250, 'SCSMD', 258-269, 'Y', 271-312, 'VQ', 315-371, 'RF', 'A'
A': 1596-2214 <SOUI>
A:Cross-references: EMBL:X05553; NID:g4760; PIDN:CAA29068.1; PID:g4761
A:Note: the authors translated the codon CAC for residue 85 as His
R:Katsoulou, C.; Tzermia, M.; Alexandraki, D.
submitted to the EMBL Data Library, May 1995
A:Description: The complete sequence of a 4.07 kb segment located on the left arm of
yeast hypothetical proteins.
A:Reference number: S55159
A:Accession: S55159
A:Molecule type: DNA
A:Residues: 175-2214 <KA2>
A:Cross-references: EMBL:X87371; NID:g854542; PIDN:CAA60825.1; PID:g854566

Matches 13; Conservative 9; Mismatches 11; Indels 1; Gaps 1;

QY 6 GFKLFEATEATSDWLNANNVPATPVAPWSQE-QGN 38
I : I I I : I I : I I I : I : I : I I

DB 950 GTTLFGTGTADFYSENGVPTQLNWDDEELGEN 983

RESULT 8
QZDPD3
pyrimidine synthesis protein PYR1-3 - slime mold (*Dictyostelium discoideum*) (fragment)
N:Contains: aspartate carbamoyltransferase (EC 2.1.3.2); carbamoyl-phosphate synthase
C:Species: *Dictyostelium discoideum*
C:Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 07-Aug-1998
C:Accession: S02800
R:Fauvre, M.; Camonis, J.H.; Jaquet, M.
Eur. J. Biochem. 179, 345-358, 1989
A:A:title: Molecular characterization of a *Dictyostelium discoideum* gene encoding a mul
A:Reference number: S02800; MID:89137111; PMID:2917570
A:Accession: S02800
A:Molecule type: DNA
A:Residues: 1-467;468-1481 <PAU>
A:CROSS-references: EMBL:X14633
A:Experimental source: strain AX3
C:Genetics:
A:Gene: PYR1-3
C:Superfamily: rudimentary enzyme; aspartate/ornithine carbamoyltransferase homology;
arabamoyl-phosphate synthase (glutamine-hydrolyzing) large chain homology; carbamoyl-p
C:Keywords: hydrolase; ligase; methyltransferase; multifunctional enzyme; pyrimidine
F:1-340/Domain: carbamoyl-phosphate synthase (ammonia) homology (fragments) <CPA>
F:1-340/Domain: carbamoyl-phosphate synthase (glutamine-hydrolyzing) small chain homo
F:158-340/Domain: trpg homology <TRG>
F:357-704/Domain: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain ho
F:367-467/Domain: biotin carboxylase homology (fragment) <BC1>
F:468-651/Domain: biotin carboxylase homology (fragment) <BC2>
F:721-1066/Domain: bacillus dihydrootetase homology <DHO>
F:1179-1477/Domain: aspartate/ornithine carbamoyltransferase homology <ACT>
F:236/Active site: Cys #status predicted

Query Match 27.5%; Score 74.5; DB 1; Length 1481;
Best Local Similarity 38.2%; Pred. No. 0.68;
Matches 13; Conservative 9; Mismatches 11; Indels 1; Gaps 1;

QY 6 GFKLFEATEATSDWLNANNVPATPVAPWSQE-QGN 38
I : I I I : I I : I I I : I : I : I I

DB 611 GTTLFGTGTADFYSENGVPTQLNWDDEELGEN 644

RESULT 9
B83605
Probable acetylpolypamine aminohydrolase PA0321 [imported] - *Pseudomonas aeruginosa* (s
C:Species: *Pseudomonas aeruginosa*
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: B83605
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.: Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:A:title: Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pa
A:Reference number: B82950; MID:20437337; PMID:10984043
A:Accession: B83605
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-344 <STO>
A:CROSS-references: GB:AE004470; GB:AB004091; MID:g9946164; PIDN:AA03710.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA0321

Query Match 25.6%; Score 69.5; DB 2; Length 344;
Best Local Similarity 37.1%; Pred. No. 0.5;
Matches 13; Conservative 7; Mismatches 14; Indels 1; Gaps 1;

Oy 1 QLHNHGFKLFTEATSDDL-NANNVPATPVAPWSQ 34


```

RESULT 2
JQ1348
carbamoyl-phosphate synthase (ammonia) (EC 6.3.4.16) precursor - human
N:Alternate names: carbamyl phosphate synthetase I; carbon-dioxide-ammonia ligase
C:Species: Homo sapiens (man)
C:Date: 30-Jun-1992 #sequence_revision 02-Jun-1994 #text_change 16-Jun-2000
C:Accession: JQ1348
R:Haraguchi, Y.; Uchino, T.; Takiguchi, M.; Endo, F.; Mori, M.; Matsuda, I.
Gene 107, 335-340, 1991
A:Title: Cloning and sequence of a cDNA encoding human carbamyl phosphate synthetase I.
A:Reference number: JQ1348; MUID:92084128; PMID:1840546
A:Accession: JQ1348
A:Molecule type: mRNA
A:Residues: 1-1500 <HAR>
A:Cross-references: DDBJ:D90282; NID:g219552; PIDN:BAAL4328.1; PID:g219553
A:Comment: This is the first enzyme of the urea cycle; it catalyzes the synthesis of carbamoyl phosphate.
C:Genetics:
A:Gene: GDB:CF51
A:Cross-references: GDB:119799; OMIM:237300
A:Map position: 2q33-q36
A:Superfamily: carbamoyl-phosphate synthase (ammonia); biotin carboxylase homology; carbamoyl-phosphate synthase (glutamine-hydrolyzing) small chain homology; trpG
C:Keywords: ATP; ligase; urea cycle
F:1-38/Domain: transit peptide (mitochondrion) #status predicted <TRP>
F:39-1500/Product: carbamoyl-phosphate synthase (ammonia) #status predicted <CAR>
F:48-1479/Domain: carbamoyl-phosphate synthase (ammonia) homology <CPA>
F:48-395/Domain: carbamoyl-phosphate synthase (glutamine-hydrolyzing) small chain homology <TRG>
F:220-395/Domain: trpG homology <TRG>
F:424-1476/Domain: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain homology <BC1>
F:424-886/Domain: biotin carboxylase homology <BC1>
F:975-1426/Domain: biotin carboxylase homology <BC2>

Query Match 98.2%; Score 266; DB 1; Length 1500;
Best Local Similarity 98.0%; Pred. No. 3.1e-25;
Matches 50; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 QLNHEGFKLFATEATSDWLNANNVPATPVAPWSQSQNPSSIRKLRDQ 51
DB 1380 QLNHEGFKLFATEATSDWLNANNVPANVPANFWSQSQNPSSIRKLRDQ 1430

RESULT 3
I51170
carbamyl phosphate synthetase - bullfrog
C:Species: Rana catesbeiana (bullfrog)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 22-Jun-1999
C:Accession: I51170
R:Helbing, C.C.; Atkinson, B.G.
J. Biol. Chem. 269, 11743-11750, 1994
A:Title: 3,5,3'-Triiodothyronine-induced carbamyl phosphate synthetase gene expression
A:Reference number: A53567; MUID:94216272; PMID:8163471
A:Accession: I51170
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1496 <HEL>
A:Cross-references: EMBL:U05193; NID:g467222; PIDN:AAAL9016.1; PID:g467223
C:Superfamily: carbamoyl-phosphate synthase (ammonia); biotin carboxylase homology; carbamoyl-phosphate synthase (glutamine-hydrolyzing) small chain homology; trpG
C:Keywords: ATP
F:43-1475/Domain: carbamoyl-phosphate synthase (ammonia) homology <CPA>
F:43-392/Domain: carbamoyl-phosphate synthase (glutamine-hydrolyzing) small chain homology <TRG>
F:216-392/Domain: trpG homology <TRG>
F:421-1472/Domain: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain homology <BC1>
F:972-1422/Domain: biotin carboxylase homology <BC2>
F:290/Active site: Cys #status predicted

Query Match 68.1%; Score 184.5; DB 2; Length 1496;
Best Local Similarity 72.0%; Pred. No. 7.1e-15;
Matches 36; Conservative 8; Mismatches 5; Indels 1; Gaps 1;

```

```

QY 2 LHNHEGFKLFATEATSDWLNANNVPATPVAPWSQSQNPSSIRKLRDQ 51
DB 1378 LKDEGFKLYATEATADWLNANDITATPVAPWSQSQS-GPSSYKLIKES 1426

RESULT 4
S53602
carbamoyl-phosphate synthase (glutamine-hydrolyzing) (EC 6.3.5.5) III precursor, hepa
N:Alternate names: carbamyl phosphate synthetase III
C:Species: Squalus acanthias (spiny dogfish)
C:Date: 28-Oct-1996 #sequence_revision 07-Feb-1997 #text_change 07-Dec-1999
C:Accession: S53602; S72160; S71062
R:Hong, J.; Salo, W.L.; Lusty, C.J.; Anderson, P.M.
J. Mol. Biol. 243, 131-140, 1994
A:Title: Carbamyl phosphate synthetase III, an evolutionary intermediate in the trans
A:Reference number: S53602; MUID:95018247; PMID:7932737
A:Accession: S53602
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-1502 <HON1>
A:Cross-references: EMBL:L31362
A:Accession: S72160
A:Molecule type: protein
A:Residues: 39-47 <HON2>
R:Anderson, P.M.
submitted to the EMBL Data Library, September 1994
A:Reference number: S71062
A:Accession: S71062
A:Molecule type: mRNA
A:Residues: 1-888, 'E', 890-1502 <AND>
A:Cross-references: EMBL:L31362; NID:g1256714; PIDN:AAA96435.1; PID:g530209
C:Genetics:
A:Genome: nuclear
C:Superfamily: carbamoyl-phosphate synthase (ammonia); biotin carboxylase homology; C
omology: carbamoyl-phosphate synthase (glutamine-hydrolyzing) small chain homology; t
C:Keywords: arginine biosynthesis; ligase; mitochondrion; pyrimidine nucleotide biosy
F:1-38/Domain: transit peptide (mitochondrion) #status predicted <TRP>
F:39-1502/Product: carbamyl-phosphate synthetase (glutamine-hydrolyzing) III #status
F:48-1480/Domain: carbamoyl-phosphate synthase (ammonia) homology <CPA>
F:48-396/Domain: carbamoyl-phosphate synthase (glutamine-hydrolyzing) small chain hom
F:220-396/Domain: trpG homology <TRG>
F:425-1477/Domain: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain h
F:425-887/Domain: biotin carboxylase homology <BC1>
F:976-1427/Domain: biotin carboxylase homology <BC2>
F:294/Active site: Cys #status predicted

Query Match 60.9%; Score 165; DB 2; Length 1502;
Best Local Similarity 56.9%; Pred. No. 2.2e-12;
Matches 29; Conservative 7; Mismatches 15; Indels 0; Gaps 0;

QY 1 QLNHEGFKLFATEATSDWLNANNVPATPVAPWSQSQNPSSIRKLRDQ 51
DB 1381 QLNHEGFKLYATEATSTWLNVDVTPAPWSPTAEHSSAPSTFKLIHDG 1431

RESULT 5
D64472
carbamoyl-phosphate synthase, pyrimidine-specific, large subunit (EC 6.3.-.-) - Metha
C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: D64472
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak
e, R.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,
son, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc
A:Reference number: A64300; MUID:96337999; PMID:8688087.
A:Accession: D64472
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-618 <BUL>
A:Cross-references: GB:U67578; GB:L77117; NID:g2826403; PIDN:AA899391.1; PID:g1592027

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2003, 02:11:17 ; Search time 23.5 Seconds
(without alignments)
208.707 Million cell updates/sec

Title: US-09-585-077C-4_COPY_1380_1430
Perfect score: 271
Sequence: 1 QHNEGKFLFATEATSDWLN.....PSQGNPSLSIRKLIRDG 51
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues 283308
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR76: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	271	100.0	1500	1 STRTCA	carbamoyl-phosphat
2	266	98.2	1500	1 JQ1348	carbamoyl-phosphat
3	184.5	68.1	1496	2 I51170	carbamoyl-phosphat
4	165	60.9	1502	2 S53602	carbamoyl-phosphat
5	82	30.3	618	2 D64472	carbamoyl-phosphat
6	79	29.2	2198	2 T20371	hypothetical prote
7	74.5	27.5	1042	2 S23738	pyr1-3 protein - s
8	74.5	27.5	1481	1 Q2D0P3	pyrimidine synthe
9	69.5	25.6	344	2 B83605	probable acetylpol
10	69.5	25.6	1115	2 A70990	carbamoyl-phosphat
11	68.5	25.3	1071	2 F39845	carbamoyl-phosphat
12	65	24.0	2214	1 Q2BVD2	pyrimidine synthe
13	64	23.6	863	2 A82722	DNA-directed RNA p
14	64	23.6	863	2 B97504	RNA polymerase [lm
15	63	23.2	2223	2 S65074	pyrimidine synthe
16	63	23.2	2244	2 T11616	carbamoyl-phosphat
17	62	22.9	336	2 G95003	membrane protein [
18	62	22.9	340	2 B87876	conserved hypothet
19	59.5	22.0	537	2 B87414	RNA polymerase sig
20	59.5	22.0	1249	2 T14150	vesicle associated
21	59.5	22.0	2225	1 A23443	pyrimidine synthe
22	58.5	21.6	277	2 AE1182	sugar ABC transpor
23	58.5	21.6	277	2 AF1539	sugar ABC transpor
24	58.5	21.6	715	2 S46228	hypothetical prote
25	58	21.4	452	2 C72277	phosphoribosylamin
26	57	21.0	2242	2 A57541	pyrimidine synthe
27	56.5	20.8	673	4 F40201	artifact-warning s
28	56.5	20.8	700	2 B82788	metalloproteinase X
29	56.5	20.8	771	2 S51421	hypothetical prote

30 56.5 20.8 1129 2 H86975
31 56 20.7 297 2 T18637
32 56 20.7 1162 2 AD3317
33 56 20.7 1524 2 G84721
34 56 20.7 1750 2 G84649
35 55 20.3 190 2 G84018
36 55 20.3 244 1 UESY25
37 55 20.3 254 2 S08511
38 55 20.3 291 2 T08848
39 55 20.3 521 2 S62794
40 55 20.3 2105 2 T18968
41 54.5 20.1 662 2 S56303
42 54.5 20.1 1062 2 H83966
43 54.5 20.1 1228 2 L40468
44 54 19.9 347 2 S55629
45 54 19.9 781 2 T41551

ALIGNMENTS

RESULT 1
SYRTCA
carbamoyl-phosphate synthase (ammonia) (EC 6.3.4.16) I precursor - rat
N:Alternate names: carbamyl phosphate synthetase I; carbon-dioxide-ammonia ligase
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 31-Dec-2000
C:Accession: A28481; A23580
R:Adage, M.; Howell, B.W.; Burak, R.; Lusty, C.J.; Shore, G.C.
J. Biol. Chem. 262, 10415-10418, 1987
A:Title: Rat carbamyl-phosphate synthetase I gene. Promoter sequence and tissue-speci
A:Reference number: A28481; MUID:87280089; PMID:3038878
A:Accession: A28481
A:Molecule type: DNA
A:Residues: 1-42 <LAG>
A:Cross-references: GB:J02805; NID:G203577; PID:AAA40959.1; PID:G203578
R:Nyuno, H.; Broglie, K.E.; Widgren, E.E.; Lusty, C.J.
J. Biol. Chem. 260, 9346-9356, 1985
A:Title: Characterization and derivation of the gene coding for mitochondrial carbam
A:Reference number: A23580; MUID:85261323; PMID:2991241
A:Accession: A23580
A:Molecule type: mRNA
A:Residues: 1-1500 <NYD>
A:Cross-references: GB:M11710
R:Potter, M.D.; Powers-Lee, S.G.
J. Biol. Chem. 267, 2023-2031, 1992
A:Title: Location of the ATP-gamma-phosphate-binding sites on rat liver carbamoyl-pho
A:Reference number: A42051; MUID:92112936; PMID:1730733
A:Contents: annotation; MgATP binding sites
C:Comment: This enzyme, which requires N-acetylglutamate as allosteric activator, cat
C:Comment: This enzyme binds two molecules of ATP at discrete sites for use in discre
C:Superfamily: carbamoyl-phosphate synthase (ammonia); biotin carboxylase homology; c
omology; carbamoyl-phosphate synthase (glutamine-hydrolyzing) small chain homology; t
C:Keywords: ATP; duplication; ligase; mitochondrion; urea cycle
F:1-38/Domain: transit peptide (mitochondrion) #status predicted <TNP>
F:39-1500/Product: carbamoyl-phosphate synthase (ammonia) #status predicted <MAT>
F:48-1479/Domain: carbamoyl-phosphate synthase (ammonia) homology <CFA>
F:48-395/Domain: trp homology <TRG>
F:424-1476/Domain: carbamoyl-phosphate synthase (glutamine-hydrolyzing) small chain h
F:424-886/Domain: biotin carboxylase homology <BCI>
F:631-638,1327-1348/Region: ATP binding (for bicarbonate activation) #status experime
F:975-1426/Domain: biotin carboxylase homology <BC2>
F:1310-1317,1445-1454/Region: ATP binding (for carbamate activation) #status experime

Query Match 100.0%; Score 271; DB 1; Length 1500;

Best Local Similarity 100.0%; Pred. No. 7.2e-26; Mismatches 0; Indels 0; Gaps 0;

Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QHNEGKFLFATEATSDWLNANNVPATVPWPSQGNPSLSIRKLIRDG 51
|||||
Db 1380 QHNEGKFLFATEATSDWLNANNVPATVPWPSQGNPSLSIRKLIRDG 1430

CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pot_sequences.

XX SQ Sequence 60 AA;

Query Match 23.2%; Score 63; DB 22; Length 60;

Best Local Similarity 29.3%; Pred. No. 0.45;

Matches 12; Conservative 8; Mismatches 17; Indels 4; Gaps 1;

QY 4 NEGFKLEATEATSDWLNANNVPATPVAMPSPQGNPSLSI 44

DB 21 NSPFKNFTGRGQAWL----MPVIPALWEAKVGRSPVRS 57

RESULT 15

ABP39849
ID ABP39849 standard; Protein; 228 AA.

XX AC ABP39849;

DT 24-JUL-2002 (first entry)

DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4694.

DE Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
KW antibacterial; gene therapy.

XX OS Staphylococcus epidermidis.

XX PN US6380370-B1.

XX PD 30-APR-2002.

XX PF 13-AUG-1998; 98US-0134001.

XX PR 14-AUG-1997; 97US-055779P.

XX PR 08-NOV-1997; 97US-064964P.

XX PA (GENO-) GENOME THERAPEUTICS CORP.

XX PI Doucette-Stamm LA, Bush D;

XX DR WPI; 2002-381255/41.

XX DR N-PSDB; ABN92394.

XX PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis
PT polypeptide, useful for diagnosing and treating bacterial infections -

XX PS Disclosure; SEQ ID 4694; 267pp; English.

XX CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP35124 to ABP3960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences
CC can also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life
CC cycle or inhibit S. epidermidis infection.

XX CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
XX USPTO web site.

XX SQ Sequence 228 AA;

Query Match 22.9%; Score 62; DB 23; Length 228;

Best Local Similarity 26.0%; Pred. No. 3.6;

Matches 13; Conservative 11; Mismatches 22; Indels 4; Gaps 1;

QY 1 QHNEGKLFATATSDWLNANNVPATPVAMPSPQGNPSLSIRKLIRD 50

Db 157 KLKEEDFDILGIEENVINWLTKSNIKSKDIS-----EGTNVEIEKIELLLKD 202

Search completed: September 13, 2003, 02:12:45
Job time : 53 secs

CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 692 AA;
 SQ
 Query Match 23.4%; Score 63.5; DB 22; Length 692;
 Best Local Similarity 37.8%; Pred. No. 9.7;
 Matches 14; Conservative 7; Mismatches 9; Indels 7; Gaps 2;

QY 1 QLNHEGFKLFATEATSDMLNANNVPATPVAVPQSEQ 37
 :||:||||: | |:: |:-| | |||
 Db 232 ELHHKGFKLWFDLIGLWVSQNLPL----QW---EQQ 261

RESULT 13
 ABG09933
 ID ABG09933 standard; Protein; 1798 AA.
 XX
 AC ABG09933;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #9924.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI: 2001-639362/73.
 DR N-PSDB; AAS74120.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 20; SEQ ID NO 40292; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving

CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1798 AA;
 SQ
 Query Match 23.4%; Score 63.5; DB 22; Length 1798;
 Best Local Similarity 37.8%; Pred. No. 34;
 Matches 14; Conservative 7; Mismatches 9; Indels 7; Gaps 2;

QY 1 QLNHEGFKLFATEATSDMLNANNVPATPVAVPQSEQ 37
 :||:||||: | |:: |:-| | |||
 Db 180 ELHHKGFKLWFDLIGLWVSQNLPL----QW---EQQ 209

RESULT 14
 AA011745
 ID AA011745 standard; Protein; 60 AA.
 XX
 AC AA011745;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 25637.
 XX
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation.
 XX
 OS Homo sapiens.
 XX
 PN WO200164835-A2.
 XX
 PD 07-SEP-2001.
 XX
 PF 26-FEB-2001; 2001WO-US04927.
 XX
 PR 28-FEB-2000; 2000US-0515126.
 PR 18-MAY-2000; 2000US-0577409.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;
 XX
 DR WPI: 2001-514838/56.
 DR N-PSDB; AAI91676.
 XX
 PT Isolated nucleic acids and polypeptides, useful for preventing
 PT diagnosing and treating e.g. leukaemia, inflammation and immune
 PT disorders -
 XX
 PS Claim 20; SEQ ID NO 25637; 1399pp + Sequence Listing; English.
 XX
 CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and

Best Local Similarity 37.0%; Pred. No. 0.31;
Matches 17; Conservative 1; Mismatches 12; Indels 16; Gaps 2;

QY 18 WLNNANVPATPVAV-----PSQEQGNPSSLKIRLIRDG 51
||| ||| ||| ||| ||| |||
DB 13 WL---TPVIPVINTKADSGIQRDPSQHGETSPSLKIEKLAGHG 54

RESULT 9
ABP75436
ID ABP75436 standard; Protein; 94 AA.
XX AC ABP75436;
DT DT
DE 10-FEB-2003 (first entry)
DE Human secretory polypeptide SPTM SEQ ID NO 620.

XX Human; SPTM; autoimmune disorder; inflammatory disorder; AIDS; anaemia;
KW asthma; Crohn's disease; neurological disorder; epilepsy; cancer;
KW Huntington's disease; Alzheimer's disease; Creutzfeldt-Jakob disease;
KW multiple sclerosis; Parkinson's disease; cell proliferative disorder;
KW anti-inflammatory; immunosuppressive; neuroprotective; neotropic;
KW neuroleptic; anticonvulsant; cytostatic; antiparkinsonian; amniolytic;
KW antipsoriatic; antianaemic; anti-HIV; human immunodeficiency virus;
KW secretory polynucleotide; secretory protein.
XX Homo sapiens.
OS
QN WQO2023876-A2.
PN XX
XX PD 24-OCT-2002.
PF XX
PE 27-MAR-2002; 2002WO-US09921.
PP XX
PR 29-MAR-2001; 2001US-280057P.
PR 29-MAR-2001; 2001US-280068P.
PR 16-MAY-2001; 2001US-291280P.
PR 17-MAY-2001; 2001US-291829P.
PR 17-MAY-2001; 2001US-291849P.
PR 19-JUN-2001; 2001US-299428P.
PR 20-JUN-2001; 2001US-299776P.
PR 20-JUN-2001; 2001US-300001P.
XX
XX (INCYTE GENOMICS INC.)
PA Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
PI Daughtrey GE, Hillman JL, Yu JY, Tusson O, Rap PE, Amshey SR;
PI Dufourcy SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;
PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;
PI Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;
XX
XX WPI: 2003-075543/07.
DR N-PSDB; ABZ35886.
DR XX

New human secretory proteins and polynucleotides, useful for
diagnosing, treating or preventing autoimmune/inflammatory disorders
(e.g. AIDS), neurological disorders (e.g. Alzheimer's), or cell
proliferations or cancers -
XX
XX Claim 27; SEQ ID NO 620; 458pp + Sequence Listing; English.

The invention relates to a secretory polynucleotide (designated sptm) comprising any of 567 polynucleotide sequences (ABZ35837-ABZ36403), a naturally occurring polynucleotide sequence at least 90 % identical to the polynucleotide sequence, a polynucleotide complementary to them or an RNA equivalent of them. The polypeptide or polynucleotide are useful for treating, preventing or diagnosing a disease or condition associated with the expression of functional SPTM. These are particularly useful for diagnosing, treating or preventing autoimmune/inflammatory disorders (e.g. acquired immunodeficiency syndrome, anaemia, asthma or Crohn's disease), neurological disorders (e.g. epilepsy, Huntington's disease, dementia, stroke, Alzheimer's disease, Creutzfeldt-Jakob disease,

[illegible]

XX Disclosure; Fig 11; 171pp; English.

PS The present invention relates to screening for susceptibility to

CC suboptimal urea cycle function or to bone marrow transplant toxicity

CC by detecting a polymorphism in the carbamyl phosphate synthase I

CC (CPSI) gene. The method is used to detect subjects at risk of

CC hepatitis, sclerosis, pulmonary hypertension and bone marrow

CC transplant toxicity. These conditions can be treated or prevented

CC by administration of a nitric oxide precursor or by gene

CC therapy (administration of sequences that encode CPSI).

XX SQ Sequence 1500 AA;

Query Match 100.0%; Score 271; DB 22; Length 1500;

Best Local Similarity 100.0%; Pred. No. 3.4e-28;

Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 QLNHEGKFLFATEATSDWLNANNVPATPVAMPQSGQNPSSLSIRKLIRG 51

DB 1380 QLNHEGKFLFATEATSDWLNANNVPATPVAMPQSGQNPSSLSIRKLIRG 1430

RESULT 2

AAAR30636

ID AAR30636 standard; Protein; 1500 AA.

AC AAR30636;

XX 06-MAY-1993 (first entry)

DT hcPSI.

DE Human; carbamyl phosphate synthetase I; CPSI; probe; mutation;

KW CPSI deficiency disease; rat.

XX Homo sapiens.

XX JP04335889-A.

XX 24-NOV-1992.

XX 09-MAY-1991; 91JP-0135902.

XX 09-MAY-1991; 91JP-0135902.

XX (HARA/) HARAGUCHI Y.

PA (MATS/) MATSUDA I.

PA (MORI/) MORI M.

XX WPI; 1993-006237/01.

DR N-PSDB; AAQ34768.

XX Carbamyl phosphate synthetase I gene - used to detect diseases

PT caused by carbamyl phosphate synthetase by using overlapped

PT cDNA contg. total human cDNA sequence

XX Disclosure; Fig 2-9; 12pp; Japanese.

XX This sequence is the human carbamyl phosphate synthetase I (CPSI)

CC protein. The DNA encoding this peptide or fragments of it may be

CC used as probes to detect mutations in the CPSI gene. CPSI deficiency

CC disease may be detected by using overlapping cDNA representing the

CC full length cDNA sequence of human CPSI. The DNA sequence was

CC isolated using three amplified fragments derived from protein coding

CC regions of the rat CPSI cDNA as probes.

XX SQ Sequence 1500 AA;

Query Match 98.2%; Score 266; DB 14; Length 1500;

Best Local Similarity 98.0%; Pred. No. 1.7e-27;

Matches 50; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 QLNHEGKFLFATEATSDWLNANNVPATPVAMPQSGQNPSSLSIRKLIRG 51

DB 1380 QLNHEGKFLFATEATSDWLNANNVPATPVAMPQSGQNPSSLSIRKLIRG 1430

RESULT 4

AAAB49224

ID AAB49224 standard; protein; 1500 AA.

XX AAB49224;

AC 13-MAR-2001 (first entry)

DT Human carbamyl phosphate synthase protein.

DE Urea cycle; carbamyl phosphate synthase I; CPSI; hepatitis;

KW sclerosis; pulmonary hypertension; bone marrow; gene therapy.

XX

OY 1 QLNHEGKFLFATEATSDWLNANNVPATPVAMPQSGQNPSSLSIRKLIRG 51

DB 1380 QLNHEGKFLFATEATSDWLNANNVPANVPAMPQSGQNPSSLSIRKLIRG 1430

RESULT 3

AAAB49222

ID AAB49222 standard; protein; 1500 AA.

XX AAB49222;

AC 13-MAR-2001 (first entry)

DT Human N1405 carbamyl phosphate synthase I protein.

DE Urea cycle; carbamyl phosphate synthase I; CPSI; hepatitis;

KW sclerosis; pulmonary hypertension; bone marrow; gene therapy.

XX Homo sapiens.

XX WO200073322-A1.

PN 07-DEC-2000.

PD 01-JUN-2000; 2000WO-US15079.

XX 01-JUN-1999; 99US-0323472.

XX (UYVA-) UNIV VANDERBILT.

XX Summar ML, Christman BW;

XX WPI; 2001-049926/06.

XX Detecting susceptibility to suboptimal urea cycle function, e.g. bone

PT marrow transplant toxicity, comprises identifying a polymorphism in the

PT gene for carbamyl phosphate synthase

XX Claim 57; Fig 12; 171pp; English.

XX The present invention relates to screening for susceptibility to

CC suboptimal urea cycle function or to bone marrow transplant toxicity

CC by detecting a polymorphism in the carbamyl phosphate synthase I

CC (CPSI) gene. The method is used to detect subjects at risk of

CC hepatitis, sclerosis, pulmonary hypertension and bone marrow

CC transplant toxicity. These conditions can be treated or prevented

CC by administration of a nitric oxide precursor or by gene

CC therapy (administration of sequences that encode CPSI).

XX SQ Sequence 1500 AA;

Query Match 98.2%; Score 266; DB 22; Length 1500;

Best Local Similarity 98.0%; Pred. No. 1.7e-27;

Matches 50; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 QLNHEGKFLFATEATSDWLNANNVPATPVAMPQSGQNPSSLSIRKLIRG 51

DB 1380 QLNHEGKFLFATEATSDWLNANNVPANVPAMPQSGQNPSSLSIRKLIRG 1430

RESULT 4

AAAB49224

ID AAB49224 standard; protein; 1500 AA.

XX AAB49224;

AC 13-MAR-2001 (first entry)

DT Human carbamyl phosphate synthase protein.

DE Urea cycle; carbamyl phosphate synthase I; CPSI; hepatitis;

KW sclerosis; pulmonary hypertension; bone marrow; gene therapy.

XX

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2003, 00:29:37 ; Search time 52 Seconds
(without alignments)
155,674 Million cell updates/sec

Title: US-09-585-077C-4_COPY_1380_1430
Perfect score: 271
Sequence: 1 QLNHEGFKLPATRSDWLN.....PSQEQNPSSIRKLIRDG 51

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :			
A_Geneseq_19Jun03:*			
1:	/SIDS1/gcgdata/geneseg/geneseq-emb1/AA1980.DAT:*		
2:	/SIDS1/gcgdata/geneseg/geneseq-emb1/AA1981.DAT:*		
3:	/SIDS1/gcgdata/geneseg/geneseq-emb1/AA1982.DAT:*		
4:	/SIDS1/gcgdata/geneseg/geneseq-emb1/AA1983.DAT:*		
5:	/SIDS1/gcgdata/geneseg/geneseq-emb1/AA1984.DAT:*		
6:	/SIDS1/gcgdata/geneseg/geneseq-emb1/AA1985.DAT:*		
7:	/SIDS1/gcgdata/geneseg/geneseq-emb1/AA1986.DAT:*		
8:	/SIDS1/gcgdata/geneseg/geneseq-emb1/AA1987.DAT:*		
9:	/SIDS1/gcgdata/geneseg/geneseq-emb1/AA1988.DAT:*		
10:	/SIDS1/gcgdata/geneseg/geneseq-emb1/AA1989.DAT:*		
11:	/SIDS1/gcgdata/geneseg/geneseq-emb1/AA1990.DAT:*		
12:	/SIDS1/gcgdata/geneseg/geneseq-emb1/AA1991.DAT:*		
13:	/SIDS1/gcgdata/geneseg/geneseq-emb1/AA1992.DAT:*		
14:	/SIDS1/gcgdata/geneseg/geneseq-emb1/AA1993.DAT:*		
15:	/SIDS1/gcgdata/geneseg/geneseq-emb1/AA1994.DAT:*		
16:	/SIDS1/gcgdata/geneseg/geneseq-emb1/AA1995.DAT:*		
17:	/SIDS1/gcgdata/geneseg/geneseq-emb1/AA1996.DAT:*		
18:	/SIDS1/gcgdata/geneseg/geneseq-emb1/AA1997.DAT:*		
19:	/SIDS1/gcgdata/geneseg/geneseq-emb1/AA1998.DAT:*		
20:	/SIDS1/gcgdata/geneseg/geneseq-emb1/AA1999.DAT:*		
21:	/SIDS1/gcgdata/geneseg/geneseq-emb1/AA2000.DAT:*		
22:	/SIDS1/gcgdata/geneseg/geneseq-emb1/AA2001.DAT:*		
23:	/SIDS1/gcgdata/geneseg/geneseq-emb1/AA2002.DAT:*		
24:	/SIDS1/gcgdata/geneseg/geneseq-emb1/AA2003.DAT:*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length DB ID	Description
1	271	100.0 1500 22 AAB49223	Human T1405 carbam
2	266	98.2 1500 14 AAR30636	hCPSI. Homo sapie
3	266	98.2 1500 22 AAB49222	Human N1405 carbam
4	266	98.2 1500 22 AAB49224	Human carbamyl pho
5	266	98.2 1500 22 AAB49225	Human carbamyl pho
6	69.5	25.6 344 22 AAU33563	Pseudomonas aerugi
7	68	25.1 43 22 AA013747	Human polypeptide
8	66	24.4 94 22 AA009946	Human polypeptide
9	64.5	23.8 94 24 ABP75436	Human secretory po

10	64	23.6	339	22	AAU03138	Streptococcus pyog
11	64	23.6	339	23	ABP28098	Streptococcus poly
12	63.5	23.4	692	22	ABG09932	Novel human diagn
13	63.5	23.4	1798	22	ABG09933	Novel human diagn
14	63	23.2	60	22	AAO11745	Human polypeptide
15	62	22.9	228	23	ABP39849	Staphylococcus epi
16	62	22.9	228	23	ABP40650	Staphylococcus epi
17	62	22.9	340	24	ABU02808	S. pneumoniae type
18	62	22.9	619	23	AAO22174	Ranoplanin biosynt
19	61	22.5	510	22	ABB62105	Drosophila melano
20	60	22.1	59	22	AAH82976	Human immune/haema
21	60	22.1	70	22	AAU18578	Human lung antigen
22	60	22.1	70	22	AAU17846	Novel human respir
23	60	22.1	3095	23	AAE20788	Rat C3b/C4b comple
24	59.5	22.0	463	22	ABE92562	Human protein sequ
25	59.5	22.0	725	20	AAV59724	Human normal ovar
26	59.5	22.0	877	22	ABE93668	Human protein sequ
27	59.5	22.0	1181	21	AAV82707	Human apoptosis re
28	59.5	22.0	1182	21	AAV82517	Human ORFX ORF281
29	59.5	22.0	1220	21	AAV82708	Human apoptosis re
30	59.5	22.0	1220	22	AAE04770	Human vesicle traf
31	59.5	22.0	1232	22	AAU32866	Novel human secret
32	59.5	22.0	1698	20	AAV31381	REV-U5i ORF1 prote
33	59	21.8	86	22	AAU29913	Novel human secret
34	59	21.8	1012	21	AAV32241	Mouse tolloid-like
35	58.5	21.6	277	23	AB347607	Listeria monocytog
36	58.5	21.6	458	22	ABG09851	Novel human diagn
37	58	21.4	63	22	AAW91439	Human immune/haema
38	58	21.4	160	22	AAU48056	Propionibacterium
39	57.5	21.2	576	23	ABP63134	lovF gene expressi
40	57.5	21.2	576	23	ABP35722	Fungal ZBC protein
41	57	21.0	55	21	AB332094	Human secreted pro
42	57	21.0	84	22	AAO10311	Human polypeptide
43	57	21.0	102	22	AAH89598	Human immune/haema
44	57	21.0	1018	21	AAE07773	Large subunit of t
45	57	21.0	1018	22	AAE50543	B. lactofermentum

ALIGNMENTS

RESULT 1	
AAB49223	
ID	AAB49223 standard; protein; 1500 AA.
XX	AAB49223;
AC	XX
DT	13-MAR-2001 (first entry)
DE	Human T1405 carbamyl phosphate synthase I DNA.
DE	XX
XX	Urea cycle; carbamyl phosphate synthase I; CPSI; hepatitis;
KW	XX
KW	XX
OS	XX
OS	XX
PN	WO200073322-A1.
XX	XX
PD	07-DEC-2000.
PF	XX
PF	XX
PR	01-JUN-2000; 2000WO-US15079.
PR	XX
PR	01-JUN-1999; 99US-0323472.
XX	XX
PA	(UYVA-) UNIV VANDERBILT.
XX	XX
PI	Summar ML, Christman BW;
XX	XX
DR	WPI; 2001-049926/06.
XX	XX
PT	Detecting susceptibility to suboptimal urea cycle function, e.g. bone
PT	marrow transplant toxicity, comprises identifying a polymorphism in the
PT	gene for carbamylphosphate synthase

Query Match 51.6%; Score 41.8; DB 10; Length 242;
Best Local Similarity 95.6%; Pred. No. 7.8e-06;
Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 37 GCCACCCCACTGGCATGGCCGCTCTCAAGAGGACAGAATCCCAGC 81
 || |||||
 db 1 GCTACCCCACTGGCATGGCCATCTCAAGAGGACAGAATCCCAGC 45

RESULTS

```

US-09-801-274-1768
; Sequence 1768, Application US/09801274
; Patent No. US20020032319A1
; GENERAL INFORMATION:
; APPLICANT: Cargill, Michele
; APPLICANT: Ireland, James S.
; APPLICANT: Lander, Eric S.
; TITLE OF INVENTION: HUMAN SINGLE NUCLEOTIDE POLYMORPHISMS
; FILE REFERENCE: 2825.2009-001
; CURRENT APPLICATION NUMBER: US/09/801,274
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 60/187,510
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: US 60/206,129
; PRIOR FILING DATE: 2000-05-22
; NUMBER OF SEQ ID NOS: 1802
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1768
; LENGTH: 31
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-801-274-1768

```

Query Match 37.8%; Score 30.6; DB 9; Length 31;
Best Local Similarity 96.8%; Pred. No. 0.083;
Matches 30: Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 26 ACAATGTCCCTGCCACCCAGTGGCATGGCC 56
db 1 ACAATGTCCCTGCCACCCAGTGGCATGGCC 31

RESULT 11

```

US-09-764-891-5507
; Sequence 5507, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: FC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; Current Filing Date: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5507
; LENGTH: 3591
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-5507

```

Query Match 36.3%; Score 29.4; DB 11; Length 3591;
Best Local Similarity 63.4%; Pred. No. 0.39;
Matches 45; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Qy 10 GACTGGCTCAAGCCACAATGTCCTGCCACCCCAGTGGCATGCGCCTCTCAAGAAGGA 69
||||||| || ||| ||| ||||| | | ||||| || | ||| |
Dβ 1746 GACTGGCTGAGCCCCCACGATGTCCTGGCGTCA CCGCGTCCTGGCGCCTCACTGGAGGA 1805

QY 70 CAGCATTCCAG 80

Db 1806 GGGACACTCAG 1816

RESUIT 12

RESUB. 12
US-10-068-674-1/c
; Sequence 1, Application US/10068674
; Publication No. US20020177203A1
; GENERAL INFORMATION:
; APPLICANT: Kevirikko, Kari
; APPLICANT: Pihlajaniemi, Taina A.
; TITLE OF INVENTION: 2 SUBUNIT OF POLYI-4-HYDROXYLASE,
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCE ENCODING SUCH SUBUNIT AND METHODS
; TITLE OF INVENTION: FOR PRODUCING THE SAME
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036

```

1  ZAP: 10030
2
3  COMPUTER READABLE FORM:
4
5  MEDIUM TYPE: Floppy disk
6
7  COMPUTER: IBM PC Compatible
8
9  OPERATING SYSTEM: PC-DOS/MS-DOS
10
11  SOFTWARE: PatentIn Release #1.0, Version #1.30
12
13  CURRENT APPLICATION DATA:
14
15  APPLICATION NUMBER: US/10/068,674
16
17  FILING DATE: 06-FEB-2002
18
19  CLASSIFICATION:
20
21  PRIOR APPLICATION DATA:
22
23  APPLICATION NUMBER: US/08/633,879
24
25  FILING DATE: 10-MAR-1996
26
27  CLASSIFICATION:
28

```

ATTORNEY/AGENCY INFORMATION:
NAME: Halluin, Albert P.
REGISTRATION NUMBER: 25,227
REFERENCE/DOCKET NUMBER: 8389-041
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-854-3660
TELEFAX: 415-854-3694
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2168 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 151..1761
US-10-068-674-1

Query Match 34.6%; Score 28; DB 13; Length 2168;
Best Local Similarity 63.2%; Pred. No. 1.2;
Matches 43; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 9 AGACTGGTCAAGCCCAACAATCTCCCTCCACCCCACTGGCATGGCGTCTCAAGAAGG 68

Dh 2005 AGTATGGTCTACGACACCATCCACAGACACCCCAAGAGCCAGGGCAGGTCTAAGTTGG 19

6
7
C
E
F
C
F
C
C
C
C

Db 1945 TCAGAGTC 1938

RECEIVED 13

RESUB 13
US-10-027-632-97735/c
; Sequence 97735, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108927.129

Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAAGCCACAAATGTCCCTGCCACCCCGAGTGGCGGTCT 60
|||||
Db 157 GCCACATCAGACTGGCTCAAGCCACAAATGTCCCTGCCACCCCGAGTGGCGGTCT 216
|||||
QY 61 CAAGAAGGACAGATCCCGC 81
|||||
Db 217 CAAGAAGGACAGATCCCGC 237
|||||

RESULT 6

US-10-101-510-553
; Sequence 553, Application US/10101510
; Publication No. US20030148295A1

; GENERAL INFORMATION:

; APPLICANT: WANG, JACKSON

; APPLICANT: WANG, YIXIN

; TITLE OF INVENTION: EXPRESSION PROFILES AND METHODS OF USE

; FILE REFERENCE: 15117.0012

; CURRENT APPLICATION NUMBER: US/10/101,510

; CURRENT FILING DATE: 2002-03-20

; PRIOR APPLICATION NUMBER: 60/276,947

; PRIOR FILING DATE: 2001-03-20

; NUMBER OF SEQ ID NOS: 805

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 553

; LENGTH: 5772

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: modified_base

; LOCATION: (5240)..(5365)

; OTHER INFORMATION: a, t, c, g, other or unknown

US-10-101-510-553

Query Match 100.0%; Score 81; DB 12; Length 5772;

Best Local Similarity 100.0%; Pred. No. 4e-20;

Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAAGCCACAAATGTCCCTGCCACCCCGAGTGGCGGTCT 60
|||||

Db 4310 GCCACATCAGACTGGCTCAAGCCACAAATGTCCCTGCCACCCCGAGTGGCGGTCT 4369
|||||

QY 61 CAAGAAGGACAGATCCCGC 81
|||||

Db 4370 CAAGAAGGACAGATCCCGC 4390
|||||

RESULT 7

US-09-880-107-1769

; Sequence 1769, Application US/09880107

; Patent No. US20020142981A1

; GENERAL INFORMATION:

; APPLICANT: Horne, Darci T.

; APPLICANT: Vockley, Joseph G.

; APPLICANT: Scherf, Uwe

; APPLICANT: Gene Logic, Inc.

; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer

; FILE REFERENCE: 44921-5028-WO

; CURRENT APPLICATION NUMBER: US/09/880,107

; CURRENT FILING DATE: 2001-06-14

; PRIOR APPLICATION NUMBER: US 60/211,379

; PRIOR FILING DATE: 2000-06-14

; PRIOR APPLICATION NUMBER: US 60/237,054

; PRIOR FILING DATE: 2000-10-02

; NUMBER OF SEQ ID NOS: 3950

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1769

; LENGTH: 5215

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: Genbank Accession No. US20020142981A1 D90282

US-09-880-107-1769

Query Match 98.0%; Score 79.4; DB 10; Length 5215;

Best Local Similarity 98.8%; Pred. No. 1.5e-19;

Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAAGCCACAAATGTCCCTGCCACCCCGAGTGGCGGTCT 60
|||||

Db 4295 GCCACATCAGACTGGCTCAAGCCACAAATGTCCCTGCCACCCCGAGTGGCGGTCT 4354
|||||

QY 61 CAAGAAGGACAGATCCCGC 81
|||||

Db 4355 CAAGAAGGACAGATCCCGC 4375
|||||

RESULT 8

US-09-960-352-12478

; Sequence 12478, Application US/09960352

; Patent No. US20020137139A1

; GENERAL INFORMATION:

; APPLICANT: Warren, Wesley C.

; APPLICANT: Tao, Nengbing

; APPLICANT: Byatt, John C.

; APPLICANT: Mathialagan, Nagappan

; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

; FILE OF INVENTION: MUSCLE AND FAT DEPOSITION

; FILE REFERENCE: 16511.006/37-21(10298)C

; CURRENT APPLICATION NUMBER: US/09/960,352

; CURRENT FILING DATE: 2001-09-24

; NUMBER OF SEQ ID NOS: 15112

; SEQ ID NO 12478

; LENGTH: 434

; TYPE: DNA

; ORGANISM: Bos taurus

; OTHER INFORMATION: Clone ID: 53-LIB34-037-Q1-E1-F2

US-09-960-352-12478

Query Match 92.1%; Score 74.6; DB 10; Length 434;

Best Local Similarity 95.1%; Pred. No. 6.8e-18;

Matches 77; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAAGCCACAAATGTCCCTGCCACCCCGAGTGGCGGTCT 60
|||||

Db 36 GCCACATCAGACTGGCTCAATGCCACACATGCTACCTGCTACCCCGAGTGGCGGTCT 95
|||||

QY 61 CAAGAAGGACAGATCCCGC 81
|||||

Db 96 CAAGAAGGACAGATCCCGC 116
|||||

RESULT 9

US-09-960-352-1818

; Sequence 1818, Application US/09960352

; Patent No. US20020137139A1

; GENERAL INFORMATION:

; APPLICANT: Warren, Wesley C.

; APPLICANT: Tao, Nengbing

; APPLICANT: Byatt, John C.

; APPLICANT: Mathialagan, Nagappan

; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

; FILE OF INVENTION: MUSCLE AND FAT DEPOSITION

; FILE REFERENCE: 16511.006/37-21(10298)C

; CURRENT APPLICATION NUMBER: US/09/960,352

; CURRENT FILING DATE: 2001-09-24

; NUMBER OF SEQ ID NOS: 15112

; SEQ ID NO 1818

; LENGTH: 242

; TYPE: DNA

; ORGANISM: Bos taurus

; OTHER INFORMATION: Clone ID: 08-LIB34-052-Q1-E1-B11

US-09-960-352-1818

```

US-09-902-941-93
; Sequence 93, Application US/0902941
; Patent No. US20020172952A1
; GENERAL INFORMATION:
; APPLICANT: Henderson, Robert A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Johnson, Jeffrey C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Warnerakis, Margarita
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: McNabb, Andria
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C17
; CURRENT APPLICATION NUMBER: US/09/902,941
; CURRENT FILING DATE: 2001-07-10
; NUMBER OF SEQ ID NOS: 2002
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 93
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-902-941-93

Query Match 100.0%; Score 81; DB 10; Length 531;
Best Local Similarity 100.0%; Pred. No. 3e-20;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTGCCACCCAGTGGCGCGTCT 60
DB 157 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTGCCACCCAGTGGCGCGTCT 60
QY 61 CAAGAAGGACAGATCCCGCAGC 81
DB 217 CAAGAAGGACAGATCCCGCAGC 237

RESULT 2
US-09-476-300-93
; Sequence 93, Application US/09476300
; Publication No. US20030125245A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C3
; CURRENT APPLICATION NUMBER: US/09/476,300
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 785
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 93
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-476-300-93

Query Match 100.0%; Score 81; DB 11; Length 531;
Best Local Similarity 100.0%; Pred. No. 3e-20;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTGCCACCCAGTGGCGCGTCT 60
DB 157 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTGCCACCCAGTGGCGCGTCT 60
QY 61 CAAGAAGGACAGATCCCGCAGC 81
DB 217 CAAGAAGGACAGATCCCGCAGC 237

RESULT 3
US-09-849-626-93
; Sequence 93, Application US/09849626
; Publication No. US20020197669A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya
; APPLICANT: Fanger, Gary
; APPLICANT: Wang, Aijun
; APPLICANT: Wang, Tongtong
; APPLICANT: Switzer, Anne
; APPLICANT: McNeill, Patricia
; APPLICANT: Clapper, Jonathan
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C16
; CURRENT APPLICATION NUMBER: US/09/849,626
; CURRENT FILING DATE: 2001-05-03
; NUMBER OF SEQ ID NOS: 1926
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 93
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-849-626-93

Query Match 100.0%; Score 81; DB 10; Length 531;
Best Local Similarity 100.0%; Pred. No. 3e-20;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTGCCACCCAGTGGCGCGTCT 60
DB 157 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTGCCACCCAGTGGCGCGTCT 60
QY 61 CAAGAAGGACAGATCCCGCAGC 81
DB 217 CAAGAAGGACAGATCCCGCAGC 237

RESULT 4
US-09-476-300-93
; Sequence 93, Application US/09476300
; Publication No. US20030125245A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C3
; CURRENT APPLICATION NUMBER: US/09/476,300
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 785
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 93
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-476-300-93

Query Match 100.0%; Score 81; DB 11; Length 531;
Best Local Similarity 100.0%; Pred. No. 3e-20;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTGCCACCCAGTGGCGCGTCT 60
DB 157 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTGCCACCCAGTGGCGCGTCT 60
QY 61 CAAGAAGGACAGATCCCGCAGC 81
DB 217 CAAGAAGGACAGATCCCGCAGC 237

RESULT 5
US-10-017-754-93
; Sequence 93, Application US/10017754
; Publication No. US20030054363A1
; GENERAL INFORMATION:
; APPLICANT: Henderson, Robert A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Johnson, Jeffrey C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Marnerakis, Margarita
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: McNabb, Andria
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C18
; CURRENT APPLICATION NUMBER: US/10/017,754
; CURRENT FILING DATE: 2001-10-29
; NUMBER OF SEQ ID NOS: 2004
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 93
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-017-754-93

Query Match 100.0%; Score 81; DB 14; Length 531;
Best Local Similarity 100.0%; Pred. No. 3e-20;
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: September 13, 2003, 00:27:37 ; Search time 151 Seconds
(without alignments)
1302.306 Million cell updates/sec

Title: US-09-585-077C-3_COPY_4300_4380
Perfect score: 81
Sequence: 1 gccacatcagctggctcaa.....aagaagacagatccagc 81

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1632420 seqs, 1213878141 residues

Total number of hits satisfying chosen parameters: 3264840

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_NA.*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09C_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	81	100.0	531	10	US-09-736-457-93
2	81	100.0	531	10	US-09-902-941-93
3	81	100.0	531	10	US-09-849-636-93
4	81	100.0	531	11	US-09-476-300-93
5	81	100.0	531	14	US-10-017-754-93
6	81	100.0	572	12	US-10-101-510-553
7	79.4	98.0	5215	10	US-09-880-107-1769
8	74.6	92.1	434	10	US-09-960-352-12478
9	41.8	51.6	242	10	US-09-960-352-1818
10	30.6	37.8	31	9	US-09-801-274-1768
11	29.4	36.3	3591	11	US-09-764-891-5507
12	28	34.6	2168	13	US-10-068-674-1
13	27.4	33.8	2193	13	US-10-027-632-97735
14	27.4	33.8	2193	13	US-10-027-632-103660
15	27.4	33.8	6471	14	US-10-207-685-78
16	27	33.3	288	10	US-09-983-965-4858

17	26.6	32.8	331	10	US-09-783-590-7357	Sequence 7357, Ap
18	26.4	32.6	969	10	US-09-280-197-13	Sequence 13, Appl
19	26.4	32.6	3276	10	US-09-280-197-4	Sequence 4, Appl
20	26.4	32.6	3276	10	US-09-423-126-8	Sequence 8, Appl
21	25.8	31.9	1281	14	US-10-156-761-2272	Sequence 2272, Ap
22	25.8	31.6	469	11	US-09-918-995-2365	Sequence 2365, Ap
23	25.6	31.4	1117	13	US-10-027-632-116661	Sequence 116661, Ap
24	25.4	31.4	1703	14	US-10-138-316-3	Sequence 3, Appl
25	25.4	31.4	113604	14	US-10-227-195A-1	Sequence 1, Appl
26	25.4	31.4	113604	14	US-10-227-195A-2	Sequence 2, Appl
27	25.4	31.1	567	11	US-09-918-995-26806	Sequence 26806, A
28	25.2	30.9	450	10	US-09-974-300-5526	Sequence 5526, Ap
29	25	30.6	284	10	US-09-833-381-417	Sequence 417, App
30	24.8	30.6	397	10	US-09-917-800A-897	Sequence 897, App
31	24.8	30.6	878	14	US-10-198-846-5229	Sequence 5229, Ap
32	24.8	30.6	957	13	US-10-027-632-163409	Sequence 163409, A
33	24.8	30.6	18746	12	US-10-017-161-2229	Sequence 2229, Ap
34	24.8	30.4	723	13	US-10-027-632-165578	Sequence 165578, A
35	24.6	30.4	375	14	US-10-106-698-3792	Sequence 3792, Ap
36	24.4	30.1	987	10	US-09-738-973-289	Sequence 289, App
37	24.4	30.1	987	10	US-09-854-133-289	Sequence 289, App
38	24.4	30.1	987	14	US-10-144-649A-289	Sequence 289, App
39	24.4	30.1	14917	11	US-09-909-567B-11	Sequence 11, Appl
40	24.4	30.1	146547	14	US-10-017-128-1	Sequence 1, Appl
41	24.4	29.9	14040	11	US-09-764-891-5478	Sequence 5478, Ap
42	24.2	29.9	14040	11	US-09-764-891-10205	Sequence 10205, A
43	24.2	29.9	14040	14	US-10-205-428-1004	Sequence 1004, Ap
44	24.2	29.9	14040	14	US-10-205-428-1004	Sequence 1004, Ap
45	24.2	29.9	28438	12	US-09-820-790-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-09-736-457-93
; Sequence 93, Application US/09736457
; Patent No. US20020168637A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedwick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; APPLICANT: Wang, Aijun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C15
; CURRENT APPLICATION NUMBER: US/09736,457
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 1864
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 93
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Homo sapien

US-09-736-457-93
US-09-736-457-93
1 GCCACATCAGCTGGCTCAAGCCACACATGCTCCCTGCCACCCAGTGGCATGCCGTCT 60
Best Local Similarity 100.0%; Pred. No. 3e+20;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Query Match 100.0%; Score 81; DB 10; Length 531;
US-09-736-457-93
157 GCCACATCAGCTGGCTCAAGCCACACATGCTCCCTGCCACCCAGTGGCATGCCGTCT 216
61 CAAGAAGGACAGATCCCGC 81
217 CAAGAAGGACAGATCCCGC 237

```
; EARLIER FILING DATE: 1998-08-17
; EARLIER APPLICATION NUMBER: 60/094,477
; EARLIER FILING DATE: 1998-07-29
; EARLIER APPLICATION NUMBER: 08/921,068
; EARLIER FILING DATE: 1997-08-29
; EARLIER APPLICATION NUMBER: 08/739,383
; EARLIER FILING DATE: 1998-10-29
; EARLIER APPLICATION NUMBER: 60/019,014
; EARLIER FILING DATE: 1995-12-22
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1703
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (193)..(579)
US-09-597-735-3

Query Match      31.4%; Score 25.4; DB 4; Length 1703;
Best Local Similarity 68.6%; Pred. NO. 5.2;
Matches 35; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY      3  CACATCAGACTGGCTCAACGCCCAACAATGTCCTCCGACCCCGCAGTGGCATG 53
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB     1631 CTCACAAACCGGCCCATCACCTCAAAAGGACCTGCCACCCCGCAGTGGCATG 1581

Search completed: September 13, 2003, 00:29:29
Job time : 51.5 secs
```



```

; APPLICANT: Christman, Brian
; TITLE OF INVENTION: HUMAN CARBAMYL PHOSPHATE SYNTHETASE I POLYMORPHISM AND DIAGNOSTIC
; TITLE OF INVENTION: THERAPEUTIC METHODS RELATING THERETO
; FILE REFERENCE: 1242/19
; CURRENT APPLICATION NUMBER: US/09/323,472A
; CURRENT FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 5761
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (124)..(4626)
US-09-323-472A-11

Query Match          98.0%; Score 79.4; DB 4; Length 5761;
Best Local Similarity 98.8%; Pred. No. 1.8e-19;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAAGGCCAACAAATGTCCTGCCACCCAGTGGCGTCT 60
DB 4300 GCCACATCAGACTGGCTCAAGGCCAACAAATGTCCTGCCACCCAGTGGCGTCT 4359

QY 61 CAAGAAGGACAGAAATCCGAGC 81
DB 4360 CAAGAAGGACAGAAATCCGAGC 4380

RESULT 7
US-09-323-472A-13
; Sequence 13, Application US/09323472A
; Patent No. 6346382
; GENERAL INFORMATION:
; APPLICANT: Summar, Marshall
; APPLICANT: Christman, Brian
; TITLE OF INVENTION: HUMAN CARBAMYL PHOSPHATE SYNTHETASE I POLYMORPHISM AND DIAGNOSTIC
; TITLE OF INVENTION: THERAPEUTIC METHODS RELATING THERETO
; FILE REFERENCE: 1242/19
; CURRENT APPLICATION NUMBER: US/09/323,472A
; CURRENT FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
; LENGTH: 5762
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (124)..(4626)
US-09-323-472A-13

Query Match          98.0%; Score 79.4; DB 4; Length 5762;
Best Local Similarity 98.8%; Pred. No. 1.8e-19;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAAGGCCAACAAATGTCCTGCCACCCAGTGGCGTCT 60
DB 4300 GCCACATCAGACTGGCTCAAGGCCAACAAATGTCCTGCCACCCAGTGGCGTCT 4359

QY 61 CAAGAAGGACAGAAATCCGAGC 81
DB 4360 CAAGAAGGACAGAAATCCGAGC 4380

RESULT 8
US-08-633-879C-1/c
; Sequence 1, Application US/08633879C
; Patent No. 5928922
; GENERAL INFORMATION:
; APPLICANT: Kivirikko, Kari I.
; APPLICANT: Pihlajaniemi, Taina
```

```

; APPLICANT: Helaakoski, Tarja I.
; APPLICANT: Annunen, Pia P.
; APPLICANT: Nissi, Ritva K.
; APPLICANT: No. 5928922elainen, Minna K.
; TITLE OF INVENTION: '2 SUBUNIT OF PROLYL-4-HYDROXYLASE
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING SUCH SUBUNIT AND
; TITLE OF INVENTION: METHODS FOR PRODUCING THE SAME
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds, LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/633,879C
; FILING DATE: 10-APR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Abrams, Samuel B.
; REGISTRATION NUMBER: 30,605
; REFERENCE/DOCKET NUMBER: 8389-0041-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2168 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 151...1761
; OTHER INFORMATION:
US-08-633-879C-1

Query Match          34.6%; Score 28; DB 2; Length 2168;
Best Local Similarity 63.2%; Pred. No. 0.63;
Matches 43; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 9 AGACTGGCTCAAGCCCAACATGTCCCTGCCACCCAGTGGCGTCTCAAGAGG 68
DB 2005 AGTCTGGTCTACCGCAGCCATGCCACAGACCCCGAGAGCGAGTGAAGCTGG 1946

QY 69 ACAGAAATC 76
DB 1945 TCAGAGTC 1938

RESULT 9
US-08-633-768A-4/c
; Sequence 4, Application US/08633768A
; Patent No. 6013504
; GENERAL INFORMATION:
; APPLICANT: YU, SHUKUN
; APPLICANT: BOJSEN, KIRSTEN
; APPLICANT: KRACH, KARSTEN
; APPLICANT: BOJKO, MAJA
; APPLICANT: NIELSEN, JOHN
; APPLICANT: MARCUSSEN, JAN
; TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE FROM
; TITLE OF INVENTION: A FUNGUS INFECTED ALGAE, ITS PURIFICATION
```

```
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C14
; CURRENT APPLICATION NUMBER: US/09/702,705
; CURRENT FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 1833
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 93
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-702-705-93

Query Match      100.0%; Score 81; DB 4; Length 531;
Best Local Similarity 100.0%; Pred. No. 2.3e-20;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTGCCACCCCAAGTGGCGCTCT 60
DB 157 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTGCCACCCCAAGTGGCGCTCT 216

QY 61 CAAGAAGGACAGAAATCCCGC 81
DB 217 CAAGAAGGACAGAAATCCCGC 237

RESULT 3
US-09-736-457-93
; Sequence 93, Application US/09736457
; Patent No. 6509448
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedwick, Tom
; APPLICANT: Carter, Darrik
; APPLICANT: Retter, Marc
; APPLICANT: Hannion, Jane
; APPLICANT: Fan, Liqun
; APPLICANT: Wang, Aijun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C15
; CURRENT APPLICATION NUMBER: US/09/736,457
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 1864
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 93
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-736-457-93

Query Match      100.0%; Score 81; DB 4; Length 531;
Best Local Similarity 100.0%; Pred. No. 2.3e-20;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTGCCACCCCAAGTGGCGCTCT 60
DB 157 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTGCCACCCCAAGTGGCGCTCT 216

QY 61 CAAGAAGGACAGAAATCCCGC 81
DB 217 CAAGAAGGACAGAAATCCCGC 237

RESULT 4
US-09-323-472A-3
; Sequence 3, Application US/09323472A
; Patent No. 6346382
; GENERAL INFORMATION:
; APPLICANT: Summar, Marshall
; TITLE OF INVENTION: HUMAN CARBAMYL PHOSPHATE SYNTHETASE I POLYMORPHISM AND DIAGN
; TITLE OF INVENTION: THERAPEUTIC METHODS RELATING THERETO
; FILE REFERENCE: 1242/19
; CURRENT APPLICATION NUMBER: US/09/323,472A
; CURRENT FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 5761
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: CDS
; LOCATION: (124)..(4626)
US-09-323-472A-3

Query Match      100.0%; Score 81; DB 4; Length 5761;
Best Local Similarity 100.0%; Pred. No. 4.8e-20;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTGCCACCCCAAGTGGCGCTCT 60
DB 4300 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTGCCACCCCAAGTGGCGCTCT 4359

QY 61 CAAGAAGGACAGAAATCCCGC 81
DB 4360 CAAGAAGGACAGAAATCCCGC 4380

RESULT 5
US-09-323-472A-1
; Sequence 1, Application US/09323472A
; Patent No. 6346382
; GENERAL INFORMATION:
; APPLICANT: Summar, Marshall
; APPLICANT: Christman, Brian
; TITLE OF INVENTION: HUMAN CARBAMYL PHOSPHATE SYNTHETASE I POLYMORPHISM AND DIAGN
; TITLE OF INVENTION: THERAPEUTIC METHODS RELATING THERETO
; FILE REFERENCE: 1242/19
; CURRENT APPLICATION NUMBER: US/09/323,472A
; CURRENT FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 5761
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: CDS
; LOCATION: (124)..(4626)
US-09-323-472A-1

Query Match      98.0%; Score 79.4; DB 4; Length 5761;
Best Local Similarity 98.8%; Pred. No. 1.8e-19;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTGCCACCCCAAGTGGCGCTCT 60
DB 4300 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTGCCACCCCAAGTGGCGCTCT 4359

QY 61 CAAGAAGGACAGAAATCCCGC 81
DB 4360 CAAGAAGGACAGAAATCCCGC 4380

RESULT 6
US-09-323-472A-11
; Sequence 11, Application US/09323472A
; Patent No. 6346382
; GENERAL INFORMATION:
; APPLICANT: Summar, Marshall
```

```
; APPLICANT: Summar, Marshall
; APPLICANT: Christman, Brian
; TITLE OF INVENTION: HUMAN CARBAMYL PHOSPHATE SYNTHETASE I POLYMORPHISM AND DIAGN
; TITLE OF INVENTION: THERAPEUTIC METHODS RELATING THERETO
; FILE REFERENCE: 1242/19
; CURRENT APPLICATION NUMBER: US/09/323,472A
; CURRENT FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 5761
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: CDS
; LOCATION: (124)..(4626)
US-09-323-472A-3

Query Match      100.0%; Score 81; DB 4; Length 5761;
Best Local Similarity 100.0%; Pred. No. 4.8e-20;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTGCCACCCCAAGTGGCGCTCT 60
DB 4300 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTGCCACCCCAAGTGGCGCTCT 4359

QY 61 CAAGAAGGACAGAAATCCCGC 81
DB 4360 CAAGAAGGACAGAAATCCCGC 4380

RESULT 5
US-09-323-472A-1
; Sequence 1, Application US/09323472A
; Patent No. 6346382
; GENERAL INFORMATION:
; APPLICANT: Summar, Marshall
; APPLICANT: Christman, Brian
; TITLE OF INVENTION: HUMAN CARBAMYL PHOSPHATE SYNTHETASE I POLYMORPHISM AND DIAGN
; TITLE OF INVENTION: THERAPEUTIC METHODS RELATING THERETO
; FILE REFERENCE: 1242/19
; CURRENT APPLICATION NUMBER: US/09/323,472A
; CURRENT FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 5761
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: CDS
; LOCATION: (124)..(4626)
US-09-323-472A-1

Query Match      98.0%; Score 79.4; DB 4; Length 5761;
Best Local Similarity 98.8%; Pred. No. 1.8e-19;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTGCCACCCCAAGTGGCGCTCT 60
DB 4300 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTGCCACCCCAAGTGGCGCTCT 4359

QY 61 CAAGAAGGACAGAAATCCCGC 81
DB 4360 CAAGAAGGACAGAAATCCCGC 4380

RESULT 6
US-09-323-472A-11
; Sequence 11, Application US/09323472A
; Patent No. 6346382
; GENERAL INFORMATION:
; APPLICANT: Summar, Marshall
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2003, 22:27:31 ; Search time 44.5 Seconds
(without alignments)
803.416 Million cell updates/sec

Title: US-09-585-077C-3_COPY_4300_4380
Perfect score: 81
Sequence: 1 gccacatcagctgctcaaa.....aagaagacagatccagc 81

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCUS_COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	81	100.0	495	4	US-09-323-472A-5
2	81	100.0	531	4	US-09-702-705-93
3	81	100.0	531	4	US-09-736-457-93
4	81	100.0	5761	4	US-09-323-472A-3
5	79.4	98.0	5761	4	US-09-323-472A-1
6	79.4	98.0	5761	4	US-09-323-472A-11
7	79.4	98.0	5762	4	US-09-323-472A-13
8	28	34.6	2168	2	US-08-633-879C-1
9	26.4	32.6	3276	3	US-08-633-768A-4
10	25.6	31.6	468	3	US-09-067-782A-6
11	25.4	31.4	1703	3	US-09-135-021-77
12	25.4	31.4	1703	3	US-09-135-020-3
13	25.4	31.4	1703	3	US-09-135-010A-3
14	25.4	31.4	1703	4	US-09-444-871-3
15	25.4	31.4	1703	4	US-09-597-735-3
16	25.4	31.4	1703	4	US-09-444-295-3
17	25.4	31.4	1703	4	US-09-597-732-3
18	25.4	31.4	1703	4	US-09-597-731-3
19	24.8	30.6	654	4	US-09-252-991A-12606
20	24.6	30.4	2265	2	US-08-940-332-1
21	24.6	30.4	4403765	3	US-09-103-840A-2
22	24.6	30.4	4411529	3	US-09-103-840A-1
23	24.4	30.1	987	4	US-09-370-838-289
24	24	29.6	432	4	US-09-252-991A-16133
25	24	29.6	438	4	US-09-252-991A-8292
26	24	29.6	1036	4	US-09-016-434-1308
27	23.6	29.1	1056	3	US-09-067-782A-1

C 28	23.6	29.1	1520	5	PCT-US93-07213-8	Sequence 8, Appli
C 29	23.6	29.1	2219	5	PCT-US93-07213-4	Sequence 4, Appli
C 30	23.6	29.1	2870	5	PCT-US93-07213-3	Sequence 3, Appli
C 31	23.6	29.1	3102	5	PCT-US93-07213-1	Sequence 1, Appli
C 32	23.6	29.1	152331	3	US-09-128-155-16	Sequence 16, Appli
C 33	23.6	28.9	519	4	US-09-844-634-19	Sequence 19, Appli
C 34	23.4	28.9	1886	4	US-09-149-476-93	Sequence 93, Appli
C 35	23.4	28.9	3683	4	US-09-844-634-3	Sequence 3, Appli
C 36	23.4	28.9	15602	4	US-09-844-634-17	Sequence 17, Appli
C 37	23.2	28.6	1900	4	US-09-608-285A-47	Sequence 47, Appli
C 38	23.2	28.6	1900	4	US-09-557-800C-47	Sequence 47, Appli
C 39	23.2	28.6	2294	4	US-09-608-285A-49	Sequence 49, Appli
C 40	23.2	28.6	2294	4	US-09-557-800C-49	Sequence 49, Appli
C 41	23.2	28.6	2371	4	US-09-608-285A-46	Sequence 46, Appli
C 42	23.2	28.6	2371	4	US-09-557-800C-46	Sequence 46, Appli
C 43	23.2	28.6	2497	4	US-09-608-285A-51	Sequence 51, Appli
C 44	23.2	28.6	2497	4	US-09-557-800C-51	Sequence 51, Appli
C 45	23.2	28.6	2693	4	US-09-608-285A-48	Sequence 48, Appli

ALIGNMENTS

RESULT 1
US-09-323-472A-5
; Sequence 5, Application US/09323472A
; Patent No. 6346382
; GENERAL INFORMATION:
; APPLICANT: Summar, Marshall
; APPLICANT: Christman, Brian
; TITLE OF INVENTION: HUMAN CARBAMYL PHOSPHATE SYNTHETASE I POLYMORPHISM AND DIAGNOSTIC METHODS RELATING THERETO
; FILE REFERENCE: 1242/19
; CURRENT APPLICATION NUMBER: US/09/323,472A
; CURRENT FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 495
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: exon
; LOCATION: (111)..(224)
; OTHER INFORMATION: n is G or A or C or T/U
US-09-323-472A-5

Query Match	100.0%	Score 81; DB 4; Length 495;
Best Local Similarity	100.0%	Pred. No. 2.3e-20;
Mismatches	81; Conservative 0; Indels 0; Gaps 0;	
QY	1	GCCACATCAGCTGCTCAAGCCACAAATGTCCTGCCACCCAGTGGCATGGCGCTCT 60
DB	126	GCCACATCAGCTGCTCAAGCCACAAATGTCCTGCCACCCAGTGGCATGGCGCTCT 185
QY	61	CAAGAAGGACAGATCCAGC 81
DB	186	CAAGAAGGACAGATCCAGC 206

RESULT 2
US-09-702-705-93
; Sequence 93, Application US/09702705
; Patent No. 6504010
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane

```

DEFINITION 602718213F1 NIH_MGC_49 Homo sapiens cDNA clone IMAGE:4858442 5',
            mRNA sequence.
ACCESSION  BG761337
VERSION    BG761337.1  GI:14071990
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 699)
            NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS   National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE     Unpublished
JOURNAL
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgabbs-re@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTF
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L1CML712 row: i column: 03
            High quality sequence stop: 699.
FEATURES   source
            Location/Qualifiers
            1..699
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:4858442"
                /tissue_type="melanotic melanoma, high MDR (cell line)"
                /lab_host="DH10B (phage-resistant)"
                /clone_lib="NIH_MGC_49"
            /note="Organ: skin; Vector: pOT7; Site:1: XhoI; Site:2:
            EcoRI; cDNA made by oligo-dt priming. Directionally cloned
            into EcoRI/XhoI sites using the following 5' adaptor:
            GGCACGAG(G). Size-selected >500bp for average insert size
            1.8kb. Library constructed by Ling Hong in the laboratory
            of Gerald M. Rubin (University of California, Berkeley)
            using ZAP-cDNA synthesis kit (Stratagene) and Superscript
            II RT (Life Technologies). Note: this is a NIH_MGC
            Library. I"
BASE COUNT  185 a 156 c 158 g 190 t
ORIGIN
Query Match 98.0%; Score 79.4; DB 12; Length 699;
Best Local Similarity 98.8%; Pred. No. 2.3e-14;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAAGCCCAACAATGTCCTCCGACCCAGTGGCATGGCCGTCT 60
    |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 517 GCCACATCAGACTGGCTCAAGCCCAACAATGTCCTCCGACCCAGTGGCATGGCCGTCT 576
    |||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 61 CAAGAAGGACAGATCCAGC 81
    |||||||||||||||||||
Db 577 CAAGAAGGACAGATCCAGC 597
    |||||||||||||||||||

RESULT 15
BG616938
LOCUS
DEFINITION 602615612F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4734387 5',
            mRNA sequence.
ACCESSION  BG616938
VERSION    BG616938.1  GI:13668309
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 706)
            NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS   National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE     Unpublished
            Contact: Robert Strausberg, Ph.D.
            Email: cgabbs-re@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTF
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L1CML1597 row: h column: 04
            High quality sequence stop: 706.
FEATURES   source
            Location/Qualifiers
            1..706
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:4734387"
                /lab_host="DH10B (T1 phage-resistant)"
                /clone_lib="NIH_MGC_76"
            /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site:1:
            SfiI (ggccattatggcc); Site:2: SfiI (ggccattatggcc); 5' and
            3' adaptors were used in cloning as follows: 5' adaptor
            sequence: 5'-CACGCCATTATGGC-3' and 3' adaptor sequence:
            5'-ATTCTAGAGCGCGGCGCGGACATG-gt(30)BN-3' (where B = A,
            C, or G and N = A, C, G, or T). Average insert size 1.85
            kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
            by PCR. This library was enriched for full-length clones
            and was constructed by Clontech Laboratories (Palo Alto,
            CA). Note: this is a NIH_MGC Library."
BASE COUNT  200 a 173 c 151 g 181 t
ORIGIN
Query Match 98.0%; Score 79.4; DB 10; Length 706;
Best Local Similarity 98.8%; Pred. No. 2.3e-14;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAAGCCCAACAATGTCCTCCGACCCAGTGGCATGGCCGTCT 60
    |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 283 GCCACATCAGACTGGCTCAAGCCCAACAATGTCCTCCGACCCAGTGGCATGGCCGTCT 342
    |||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 61 CAAGAAGGACAGATCCAGC 81
    |||||||||||||||||||
Db 343 CAAGAAGGACAGATCCAGC 363
    |||||||||||||||||||

Search completed: September 13, 2003, 00:27:27
Job time : 1961 secs

```

```

JOURNAL
COMMENT   Unpublished
            Contact: Robert Strausberg, Ph.D.
            Email: cgabbs-re@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L1CML1597 row: h column: 04
            High quality sequence stop: 706.
FEATURES   source
            Location/Qualifiers
            1..706
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:4734387"
                /lab_host="DH10B (T1 phage-resistant)"
                /clone_lib="NIH_MGC_76"
            /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site:1:
            SfiI (ggccattatggcc); Site:2: SfiI (ggccattatggcc); 5' and
            3' adaptors were used in cloning as follows: 5' adaptor
            sequence: 5'-CACGCCATTATGGC-3' and 3' adaptor sequence:
            5'-ATTCTAGAGCGCGGCGGCGGACATG-gt(30)BN-3' (where B = A,
            C, or G and N = A, C, G, or T). Average insert size 1.85
            kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
            by PCR. This library was enriched for full-length clones
            and was constructed by Clontech Laboratories (Palo Alto,
            CA). Note: this is a NIH_MGC Library."
BASE COUNT  200 a 173 c 151 g 181 t
ORIGIN
Query Match 98.0%; Score 79.4; DB 10; Length 706;
Best Local Similarity 98.8%; Pred. No. 2.3e-14;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAAGCCCAACAATGTCCTCCGACCCAGTGGCATGGCCGTCT 60
    |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 283 GCCACATCAGACTGGCTCAAGCCCAACAATGTCCTCCGACCCAGTGGCATGGCCGTCT 342
    |||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 61 CAAGAAGGACAGATCCAGC 81
    |||||||||||||||||||
Db 343 CAAGAAGGACAGATCCAGC 363
    |||||||||||||||||||

Search completed: September 13, 2003, 00:27:27
Job time : 1961 secs

```

Best Local Similarity 98.8%; Pred.No. 1.8e-14;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAACGCCCAACAATGTCCTGCCACCCCAAGGCGATGGCCGCTCT 60
|||||
Db 271 GCCACATCAGACTGGCTCAACGCCCAACAATGTCCTGTGCACCCCAAGTGGCGCTCT 330
|||||

QY 61 CAAGAAGGACAGATCCCGAGC 81
|||||
Db 331 CAAGAAGGACAGATCCCGAGC 351
|||||

RESULT 13
AV653332
LOCUS
DEFINITION AV653332 GLC Homo sapiens cDNA clone GLCJUG10 3', mRNA sequence.
ACCESSION AV653332
VERSION AV653332.1 GI:9874346
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 626)
AUTHORS Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
Hu,G., Gu,J., Chen,Z. and Han,Z.
TITLE Insight into hepatocellular carcinogenesis at transcriptome level
by comparing gene expression profiles of hepatocellular carcinoma
with those of corresponding noncancerous liver
Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
21625106
PUBMED 11752456
COMMENT Contact: Zequang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
Location/Qualifiers
1. 626
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="GLCJUG10"
/tissue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/clone_lib="GLC"
/notes="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 156 a 152 c 142 g 175 t 1 others
ORIGIN

Query Match 98.0%; Score 79.4; DB 9; Length 626;
Best Local Similarity 98.8%; Pred. No. 2.2e-14;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAACGCCCAACAATGTCCTGCCACCCCAAGTGGCGATGGCCGCTCT 60
|||||
Db 468 GCCACATCAGACTGGCTCAACGCCCAACAATGTCCTGTGCACCCCAAGTGGCGCTCT 527
|||||

QY 61 CAAGAAGGACAGATCCCGAGC 81
|||||
Db 528 CAAGAAGGACAGATCCCGAGC 548
|||||

RESULT 14
Bg761337
LOCUS
DEFINITION Bg761337 GLC Homo sapiens cDNA clone Bg761337 3', mRNA sequence.
ACCESSION Bg761337
VERSION Bg761337.1 GI:9874346
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 626)
AUTHORS Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
Hu,G., Gu,J., Chen,Z. and Han,Z.
TITLE Insight into hepatocellular carcinogenesis at transcriptome level
by comparing gene expression profiles of hepatocellular carcinoma
with those of corresponding noncancerous liver
Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
21625106
PUBMED 11752456
COMMENT Contact: Zequang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
Location/Qualifiers
1. 626
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="Bg761337"
/tissue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/clone_lib="GLC"
/notes="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

LOCUS AL046242 646 bp mRNA linear EST 29-FEB-2000
 DEFINITION DKFZp434E087_s1_434 (synonym: htes3) Homo sapiens cDNA clone
 ACCESSION DKFZp434E087 3', mRNA sequence.
 VERSION AL046242
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 646)
 AUTHORS Koehrer, K., Beyer, A., Mewes, H.W., Gassenhuber, J. and Wiemann, S.
 TITLE EST (Koehrer, et al.)
 JOURNAL Unpublished
 COMMENT Contact: Koehrer K
 MIPS
 Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
 This is the 3' sequence of the clone insert
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
 sequenced by BMFZ (Biomedical Research Center at the Charite,
 Berlin/Germany) within the cDNA sequencing consortium of the German
 Genome Project.
 rl sequence also available.
 This clone (DKFZp434E087) is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
 FEATURES
 Location/Qualifiers
 source
 1..646
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="DKFZp434E087"
 /tissue_type="testis"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_lib="434 (synonym: htes3)"
 /note="Vector: pSport1; Site_1: NotI; Site_2: SalI"
 BASE COUNT 171 a 128 c 156 g 190 t 1 others
 ORIGIN
 Query Match 100.0%; Score 81; DB 9; Length 646;
 Best Local Similarity 100.0%; Pred. No. 7e-15;
 Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GCCACATCAGCTGCTCAAGCCCAACATGTCCTGCCACCCAGTGGCATGCCGTCT 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 469 GCCACATCAGCTGCTCAAGCCCAACATGTCCTGCCACCCAGTGGCATGCCGTCT 410
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 61 CAAGAAGGACAGATCCCGC 81
 ||||||||||||||||||||||||
 Db 409 CAAGAAGGACAGATCCCGC 389
 ||||||||||||||||||||||||
 RESULT 10
 BG533520 949 bp mRNA linear EST 03-APR-2001
 LOCUS 601860935F2 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4071944 5',
 DEFINITION mRNA sequence.
 ACCESSION BG533520
 VERSION BG533520.1 GI:13525060
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 949)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: gcrabs@rmail.nih.gov
 Tissue Procurement: CLONETECH Laboratories, Inc.

cDNA Library Preparation: CLONETECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLCM918 row: f column: 09
 High quality sequence stop: 740.
 Location/Qualifiers
 source
 1..949
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4071944"
 /lab_host="DH10B (TI phage-resistant)"
 /clone_lib="NIH_MGC_76"
 /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
 SfiI (ggcccttcggcc); Site_2: SfiI (ggccattatggcc); 5' and
 3' adaptors were used in cloning as follows: 5' adaptor
 sequence: 5'-CAGGCCATTATGGCC-3' and 3' adaptor sequence:
 5'-ATTCATAGAGCCGCGAGCGCCGACATG-dt(30)BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size 1.85
 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
 by PCR. This library was enriched for full-length clones
 and was constructed by Clontech Laboratories (Palo Alto,
 CA). Note: this is a NIH_MGC Library."
 BASE COUNT 271 a 224 c 181 g 272 t 1 others
 ORIGIN
 Query Match 100.0%; Score 81; DB 10; Length 949;
 Best Local Similarity 100.0%; Pred. No. 8e-15;
 Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GCCACATCAGCTGCTCAAGCCCAACATGTCCTGCCACCCAGTGGCATGCCGTCT 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 181 GCCACATCAGCTGCTCAAGCCCAACATGTCCTGCCACCCAGTGGCATGCCGTCT 240
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 61 CAAGAAGGACAGATCCCGC 81
 ||||||||||||||||||||||||
 Db 241 CAAGAAGGACAGATCCCGC 261
 ||||||||||||||||||||||||
 RESULT 11
 AV656423 362 bp mRNA linear EST 16-JAN-2002
 LOCUS AV656423 GLC Homo sapiens cDNA clone GLCERB10 3', mRNA sequence.
 DEFINITION AV656423
 ACCESSION AV656423
 VERSION AV656423.1 GI:9877437
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 362)
 AUTHORS Xu, X., Huang, J., Xu, Z., Qian, B., Zhu, Z., Yan, Q., Cai, T., Zhang, X.,
 Xiao, H., Qu, J., Liu, F., Huang, Q., Cheng, Z., Li, N., Du, J., Hu, W.,
 Shen, K., Lu, G., Zhong, M., Xu, S., Gu, W., Huang, W., Zhao, X.,
 Hu, G., Gu, J., Chen, Z. and Han, Z.
 TITLE Insight into hepatocellular carcinogenesis at transcriptome level
 by comparing gene expression profiles of hepatocellular carcinoma
 with those of corresponding noncancerous liver
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
 MEDLINE 21625106
 PUBLISHED 11752456
 COMMENT Contact: Zeguang Han
 Chinese National Human Genome Center at Shanghai
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
 201203, P. R. China
 Tel: 86-21-50801919(ex.45)
 Fax: 86-21-50801922
 Email: hanzg@chgc.sh.cn
 This clone is available at CHGC in Shanghai.
 Location/Qualifiers
 FEATURES

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="L17N670205-7-B10"
/sex="F"
/lab_host="Top10F"
/clone_lib="L17N670205"
/organism="Homo sapiens"
/site_1="EcoRI"
/site_2="NotI"
The library was contributed by the Soares laboratory and it was constructed as described by Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996). Genome Research 6(9): 791-806. RNA was prepared from harvested cell culture.
BASE COUNT      178 a  159 c  134 g  154 t
ORIGIN
Query Match      100.0%; Score 81; DB 14; Length 625;
Best Local Similarity 100.0%; Pred. No. 7e-15;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTCCACCCAGTGCATGGCGCTCT 60
        |||||||
Db      221 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTCCACCCAGTGCATGGCGCTCT 280
QY      61 CAAGAAGGACAGATCCCGC 81
        |||||||
Db      281 CAAGAAGGACAGATCCCGC 301

RESULT 7
LOCUS      AV654671      636 bp      mRNA      linear      EST 15-JAN-2002
DEFINITION AV654671 GLC Homo sapiens cDNA clone GLC09 3', mRNA sequence.
ACCESSION  AV654671
VERSION     AV654671.1 GI:9875685
KEYWORDS   EST.
SOURCE      Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Xiao, H., Huang, J., Xu, Z., Qian, B., Zhu, Z., Yan, Q., Cai, T., Zhang, X.,
Shen, K., Lu, G., Fu, G., Zhong, M., Xu, S., Gu, W., Huang, W., Zhao, X.,
Hu, G., Gu, J., Chen, Z. and Han, Z.
Insight into hepatocellular carcinogenesis at transcriptome level
by comparing gene expression profiles of hepatocellular carcinoma
with those of corresponding noncancerous liver
Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
21625106
11752456
Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES
source
1..636
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="GLC09"
/tissue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/clone_lib="GLC"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT      179 a  148 c  122 g  185 t      2 others
ORIGIN

```

```

Query Match      100.0%; Score 81; DB 9; Length 636;
Best Local Similarity 100.0%; Pred. No. 7e-15;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTCCACCCAGTGCATGGCGCTCT 60
        |||||||
Db      59 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTCCACCCAGTGCATGGCGCTCT 118
QY      61 CAAGAAGGACAGATCCCGC 81
        |||||||
Db      119 CAAGAAGGACAGATCCCGC 139

RESULT 8
LOCUS      AV683932      637 bp      mRNA      linear      EST 16-JAN-2002
DEFINITION AV683932 GKC Homo sapiens cDNA clone GKCDID08 5', mRNA sequence.
ACCESSION  AV683932
VERSION     AV683932.1 GI:10285795
KEYWORDS   EST.
SOURCE      Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Xiao, H., Huang, J., Xu, Z., Qian, B., Zhu, Z., Yan, Q., Cai, T., Zhang, X.,
Shen, K., Lu, G., Fu, G., Zhong, M., Xu, S., Gu, W., Huang, W., Zhao, X.,
Hu, G., Gu, J., Chen, Z. and Han, Z.
Insight into hepatocellular carcinogenesis at transcriptome level
by comparing gene expression profiles of hepatocellular carcinoma
with those of corresponding noncancerous liver
Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
21625106
11752456
Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES
source
1..637
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="GKCDID08"
/tissue_type="hepatocellular carcinoma"
/dev_stage="Adult"
/lab_host="SOLR"
/clone_lib="GKC"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT      173 a  147 c  118 g  196 t      3 others
ORIGIN
Query Match      100.0%; Score 81; DB 9; Length 637;
Best Local Similarity 100.0%; Pred. No. 7e-15;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTCCACCCAGTGCATGGCGCTCT 60
        |||||||
Db      5 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTCCACCCAGTGCATGGCGCTCT 64
QY      61 CAAGAAGGACAGATCCCGC 81
        |||||||
Db      65 CAAGAAGGACAGATCCCGC 85

RESULT 9
AL046242/c

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1 (bases 1 to 540)
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.

21C Frontier Korean EST Project 2001

TITLE
JOURNAL
COMMENT

Unpublished
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4470

Email: yongsung@mail.kribb.re.kr

Plate: 14 row: F column: 07

High quality sequence stop: 540.

FEATURES
source

1..540
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="S20T665307-14-F07"
/sex="M"
/lab_host="Top10F"
/clone_lib="S20T665307"

/note="Organ: Stomach; Vector: pCNS; Site_1: EcoRI;
Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tabacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including EcoR
I site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized from oligo dt-selected mRNA by
priming with dt-tailed vector. The dt-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okazama-Berg method. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10F by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."

BASE COUNT 147 a 133 c 121 g 139 t

Query Match 100.0%; Score 81; DB 12; Length 540;

Best Local Similarity 100.0%; Pred. No. 6.6e-15;

Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAACGCCACAATGTCCTGCCACCCAGTGGCATGGCGTCT 60
|||||
DB 329 GCCACATCAGACTGGCTCAACGCCACAATGTCCTGCCACCCAGTGGCATGGCGTCT 389
|||||

QY 61 CAAGAGGACAGATCCCGAGC 81
|||||

DB 389 CAAGAGGACAGATCCCGAGC 409
|||||

RESULT 5

LOCUS

AV661502 564 bp mRNA linear EST 16-JAN-2002

DEFINITION AV661502 GLC Homo sapiens cDNA clone GLCGB04 3', mRNA sequence.

ACCESSION AV661502

VERSION AV661502.1 GI:9882516

KEYWORDS

SOURCE

Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 564)
Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: Zeguang Han

Chinese National Human Genome Center at Shanghai

351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai

201203, P. R. China

Tel: 86-21-50801919(ex.45)

Fax: 86-21-50801922

Email: hanzg@sh.cn

This clone is available at CHGC in Shanghai.

FEATURES
source

Location/Qualifiers

1..564

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="GLCGB04"

/tissue_type="corresponding non cancerous liver tissue"

/dev_stage="Adult"

/lab_host="SOLR"

/clone_lib="GLC"

/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:

XhoI"

BASE COUNT 140 a 140 c 129 g 155 t

Query Match

Best Local Similarity 100.0%; Score 81; DB 9; Length 564;

Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAACGCCACAATGTCCTGCCACCCAGTGGCATGGCGTCT 60
|||||
DB 453 GCCACATCAGACTGGCTCAACGCCACAATGTCCTGCCACCCAGTGGCATGGCGTCT 512
|||||

QY 61 CAAGAGGACAGATCCCGAGC 81
|||||

DB 513 CAAGAGGACAGATCCCGAGC 533
|||||

RESULT 6

LOCUS

CB153827 625 bp mRNA linear EST 29-JAN-2003

DEFINITION K-EST0211442 L17N670205 Homo sapiens cDNA clone L17N670205-7-B10

5', mRNA sequence.

ACCESSION CB153827

VERSION CB153827.1 GI:28138823

KEYWORDS

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 625)

Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,

Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and

Kim,Y.S.

21C Frontier Korean EST Project 2001

Unpublished

Contact: Kim YS

Genome Research Center

Korea Research Institute of Bioscience & Biotechnology

52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea

Tel: +82-42-860-4470

Fax: +82-42-860-4409

Email: yongsung@mail.kribb.re.kr

Plate: 7 row: B column: 10

High quality sequence stop: 625.

Location/Qualifiers

1..625

/organism="Homo sapiens"

Fax: 86-21-50801922
 Email: hanzg@chgc.sh.cn
 This clone is available at CHGC in Shanghai.

FEATURES
 source
 Location/Qualifiers
 1..366
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="GKBCUC07"
 /tissue_type="hepatocellular carcinoma"
 /dev_stage="Adult"
 /lab_host="SOLR"
 /clone_lib="GKC"
 /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 92 a 98 c 82 g 93 t 1 others
 ORIGIN

Query Match 100.0%; Score 81; DB 9; Length 366;
 Best Local Similarity 100.0%; Pred. No. 5.8e-15;
 Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAACGCCACCAACATGTCCTGCCACCCCGAGTGGCGTCT 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 193 GCCACATCAGACTGGCTCAACGCCACCAACATGTCCTGCCACCCCGAGTGGCGTCT 252
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 61 CAAGAAGGACAGATCCCGCAGC 81
 |||||||||||||||||||||||||||
 Db 253 CAAGAAGGACAGATCCCGCAGC 273
 |||||||||||||||||||||||||||

RESULT 2
 AV661012
 LOCUS AV661012 GLC Homo sapiens cDNA clone GLCGNF11 374 bp mRNA linear EST 16-JAN-2002
 DEFINITION AV661012
 ACCESSION AV661012
 VERSION AV661012.1 GI:9882026
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Xu, X., Huang, J., Xu, Z., Qian, B., Zhu, Z., Yan, Q., Cai, T., Zhang, X., Xiao, H., Qu, J., Liu, F., Huang, Q., Cheng, Z., Li, N., Du, J., Hu, W., Shen, K., Lu, G., Fu, G., Zhong, M., Xu, S., Gu, W., Huang, W., Zhao, X., Hu, G., Gu, J., Chen, Z. and Han, Z.
 Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver
 Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
 JOURNAL 21625106
 MEDLINE 11752456
 PUBMED
 CONTACT: Zeguang Han
 Chinese National Human Genome Center at Shanghai
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China
 Tel: 86-21-50801919(ex.45)
 Fax: 86-21-50801922
 Email: hanzg@chgc.sh.cn
 This clone is available at CHGC in Shanghai.
 Location/Qualifiers
 1..374
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="GLCGNF11"
 /tissue_type="corresponding non cancerous liver tissue"
 /dev_stage="Adult"
 /lab_host="SOLR"
 /clone_lib="GLC"
 /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 91 a 103 c 87 g 93 t
 ORIGIN

Query Match 100.0%; Score 81; DB 9; Length 374;
 Best Local Similarity 100.0%; Pred. No. 5.8e-15;
 Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAACGCCACCAACATGTCCTGCCACCCCGAGTGGCGTCT 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 277 GCCACATCAGACTGGCTCAACGCCACCAACATGTCCTGCCACCCCGAGTGGCGTCT 336
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 61 CAAGAAGGACAGATCCCGCAGC 81
 |||||||||||||||||||||||||||
 Db 337 CAAGAAGGACAGATCCCGCAGC 357
 |||||||||||||||||||||||||||

RESULT 3
 A1174790
 LOCUS A1174790 501 bp mRNA linear EST 11-NOV-1999
 DEFINITION HA2511 Human fetal liver cDNA library Homo sapiens cDNA, mRNA sequence.
 ACCESSION A1174790
 VERSION A1174790.1 GI:6361174
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Yu, Y., Zhang, C., Luo, L., Qiyang, S., Li, W., Wu, J., Zhou, S., Liu, M. and He, F.
 Expression profile analysis of a human fetal liver cDNA library
 Unpublished
 JOURNAL
 CONTACT: Chenggang Zhang
 Beijing Institute of Radiation Medicine
 27 Taiping Road, Beijing 100850, P.R.China
 Email: zhang.chenggang@hotmail.com.
 Location/Qualifiers
 1..501
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /tissue_type="liver"
 /dev_stage="fetal"
 /lab_host="M1061/P3"
 /clone_lib="Human fetal liver cDNA library"
 /note="Vector: pCDNA1"

BASE COUNT 143 a 124 c 110 g 124 t
 ORIGIN

Query Match 100.0%; Score 81; DB 9; Length 501;
 Best Local Similarity 100.0%; Pred. No. 6.4e-15;
 Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAACGCCACCAACATGTCCTGCCACCCCGAGTGGCGTCT 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 181 GCCACATCAGACTGGCTCAACGCCACCAACATGTCCTGCCACCCCGAGTGGCGTCT 240
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 61 CAAGAAGGACAGATCCCGCAGC 81
 |||||||||||||||||||||||||||
 Db 241 CAAGAAGGACAGATCCCGCAGC 261
 |||||||||||||||||||||||||||

RESULT 4
 BM821786
 LOCUS BM821786 540 bp mRNA linear EST 06-MAR-2002
 DEFINITION K-EST0091028 S20T665307 Homo sapiens cDNA clone S20T665307-14-F07 5', mRNA sequence.
 ACCESSION BM821786
 VERSION BM821786.1 GI:19178199
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2003, 22:26:16 ; Search time 1960 seconds
(without alignments)
1004.420 Million cell updates/sec

Title: US-09-585-077c-3_COPY_4300_4380

Perfect score: 81

Sequence: 1 gccacatcaactgctcaa.....aagaagacagaatccacgc 81

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:**

2: em_esthum:**

3: em_estin:**

4: em_estmu:**

5: em_estov:**

6: em_estpl:**

7: em_estro:**

8: em_htc:**

9: gb_est1:**

10: gb_est2:**

11: gb_htc:**

12: gb_est3:**

13: gb_est4:**

14: gb_est5:**

15: em_estfun:**

16: em_estom:**

17: em_gss_hum:**

18: em_gss_inv:**

19: em_gss_pln:**

20: em_gss_vrt:**

21: em_gss_fun:**

22: em_gss_man:**

23: em_gss_mus:**

24: em_gss_pro:**

25: em_gss_rod:**

26: em_gss_phg:**

27: em_gss_vrl:**

28: gb_gss1:**

29: gb_gss2:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query %	Score	Match	Length	DB	ID	Description
1	81	100.0	366	9	AV696724	AV696724	AV696724 AV696724
2	81	100.0	374	9	AV661012	AV661012	AV661012 AV661012
3	81	100.0	501	9	AI174790	HA2511	HA2511 Hu
4	81	100.0	540	12	BM821786	K-EST0091	BM821786 K-EST0091

5	81	100.0	564	9	AV661502	AV661502	AV661502 AV661502
6	81	100.0	625	14	CB153827	CB153827	CB153827 K-EST0211
7	81	100.0	636	9	AV654671	AV654671	AV654671 AV654671
8	81	100.0	637	9	AV683932	AV683932	AV683932 AV683932
c	81	100.0	649	9	AL046242	DFZP434E	AL046242 DFZP434E
10	81	100.0	949	10	BG533520	BG533520	BG533520 601860935
11	80	98.8	362	9	AV656423	AV656423	AV656423 AV656423
12	79.4	98.0	365	9	AV660978	AV660978	AV660978 AV660978
13	79.4	98.0	626	9	AV653332	AV653332	AV653332 AV653332
14	79.4	98.0	699	12	BG761337	BG761337	BG761337 602718213
15	79.4	98.0	706	10	BG616938	BG616938	BG616938 602615612
16	79.4	98.0	843	10	BE9711350	BE9711350	BE9711350 601651514
17	78.4	96.8	490	9	AI065054	HA0887	HA0887 Hu
c	76.2	94.1	444	9	AA833205	aml16h02.s	AA833205 aml16h02.s
19	76.2	94.1	830	10	BG569284	BG569284	BG569284 602588576
20	74.8	92.3	405	14	TS9186	Y50608.r1	TS9186 Y50608.r1
c	74.6	92.1	350	14	CB781273	AMGNNUC:C	CB781273 AMGNNUC:C
22	74.6	92.1	554	12	BM432173	1JEJ1C10.	BM432173 1JEJ1C10.
23	74.6	92.1	688	14	CB462893	CB462893	CB462893 723232 MA
24	71.4	88.1	473	9	AI786068	u157h08.y	AI786068 u157h08.y
25	71.4	88.1	492	9	AA250015	mz59409.i	AA250015 mz59409.i
26	71.4	88.1	524	9	AA288939	v001g03.i	AA288939 v001g03.i
27	71.4	88.1	629	9	AA986202	uc73e03.y	AA986202 uc73e03.y
28	71.4	88.1	706	14	CB950475	AGENCOURT	CB950475 AGENCOURT
29	71.4	88.1	740	14	CB948817	CB948817	CB948817 AGENCOURT
30	71.4	88.1	791	12	BI217440	BI217440	BI217440 602933642
31	71.4	88.1	920	10	BF533080	602073616	BF533080 602073616
32	71.4	88.1	1057	14	W29382	mm99c06.r1	W29382 mm99c06.r1
33	71.4	88.1	3161	11	AK028683	AK028683	AK028683 Mus muscu
34	70.6	87.2	307	14	T77346	Y472e05.r1	T77346 Y472e05.r1
35	70.4	86.9	725	12	BI247566	602960022	BI247566 602960022
36	69.8	86.2	511	9	AA511517	v128c06.r	AA511517 v128c06.r
37	69	85.2	541	9	AA333420	z75f02.i	AA333420 z75f02.i
c	69	85.2	553	9	AA396737	z75f02.s	AA396737 z75f02.s
39	68.6	84.7	512	12	BQ011349	BQ011349	BQ011349 UI-1-BCIP
40	67.4	83.2	521	9	AA190971	AA190971	AA190971 zp83h12.i
41	67	82.7	391	9	AV655328	AV655328	AV655328 AV655328
42	65.6	81.0	622	9	AI047399	ud65g10.y	AI047399 ud65g10.y
43	62	76.5	564	14	CB120338	CB120338	CB120338 K-EST0167
44	61.6	76.0	372	10	BF547163	BF547163	BF547163 UI-R-CI-1
45	60	74.1	745	14	CB594152	CB594152	CB594152 AGENCOURT

ALIGNMENTS

RESULT 1
AV696724
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

AV696724 366 bp mRNA linear EST 16-JAN-2002
AV696724 GRC Homo sapiens cDNA clone GKCBUC07 5', mRNA sequence.
AV696724.1 GI:10298587
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 366)
Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
Hu,G., Gu,J., Chen,Z. and Han,Z.
Insight into hepatocellular carcinogenesis at transcriptome level
by comparing gene expression profiles of hepatocellular carcinoma
with those of corresponding noncancerous liver
Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)

Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)

QY 61 CRAGAAGGACAGAAATCCGAGC 81
Db |||||||||||||||||||
96 CRAGAAGGACAGAAATCCGAGC 116

Search completed: September 12, 2003, 22:34:27
Job time : 203.5 secs

XX DE Toxicologically relevant human nucleotide sequence #1900.
XX KW Toxicologically relevant gene; toxicological response; gene; ss.
XX OS Homo sapiens.
XX PN WO2003016500-A2.
XX PD 27-FEB-2003.
XX XX 16-AUG-2002; 2002WO-US26514.
XX PF 16-AUG-2001; 2001US-313080P.
XX PR (PHAS-) PHASE-1 MOLECULAR TOXICOLOGY INC.
XX PA Neft RE, Dunn RT, Adkins K, Pickett GG, Kier LD, Schmeiser K;
XX PI Alen P;
XX DR WPI; 2003-268322/26.
XX XX Determining a toxicological response to an agent, useful for screening
PT of drugs, comprises comparing the expression profile of one or more
PT human toxic response genes to a reference gene expression profile
PT indicative of toxicity -
XX PS Claim 1; Page 447; 455pp; English.
XX CC The present invention describes a method (M1) for determining a
CC toxicological response to an agent, which comprises comparing the
CC expression profile of one or more human toxic response genes to a
CC reference gene expression profile indicative of toxicity, and so
CC determining the presence of a toxic response to the agent. Also
CC described: (1) an array comprising one or more polynucleotides selected
CC from the genes corresponding to the partial sequences given in AB282842
CC to AB284764, or their fragments of at least 20 nucleotides, or
CC homologues; and (2) determining if a gene putatively identified to be a
CC toxic response gene plays a role on toxic response pathways by
CC determining the expression profile of the gene after exposure of cells
CC or a human subject to a known toxic pharmaceutical or industrial agent,
CC comprising: (a) exposing cells to an agent or isolating cells from a
CC human subject who was exposed to an agent; (b) obtaining the test gene
CC expression profile for a putatively identified toxic response gene after
CC exposure to a known toxic pharmaceutical or industrial agent; and
CC (c) comparing the test profile to the expression profile of a gene with
CC a similar function or comparing the test profile to the expression
CC profile of that gene after exposure to other known toxic compounds. The
CC methods are useful for predicting and determining toxicological responses
CC on a cellular, organ or system level. The arrays comprising the human
CC genes are useful for toxicological screening of drugs, pharmaceutical
CC compounds and chemicals.
XX SQ Sequence 715 BP; 196 A; 173 C; 141 G; 205 T; 0 other;
Query Match 94.8%; Score 76.8; DB 25; Length 715;
Best Local Similarity 97.5%; Pred. No. 4.5e-17;
Matches 78; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 CCACATCAGATGCTCAAGCCACACATGTCCTGCGACCCAGTGGCATGGCGCTC 61
DB 49 CCTTATCAGATGCTCAAGCCACACATGTCCTGCGACCCAGTGGCATGGCGCTC 108
QY 62 AAGAAGGACAGATCCCGAGC 81
DB 109 AAGAAGGACAGATCCCGAGC 128
RESULT 15
ABX47313
ID ABX47313 standard; CDNA; 434 BP.
XX AC
ABX47313;

XX DT 21-FEB-2003 (first entry)
XX DE Bovine EST associated with lactation/muscle/fat deposition #12478.
XX KW Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
KW muscle deposition; fat deposition; genome mapping; gene identification;
KW gene analysis; cattle breeding.
XX OS Bos Taurus.
XX PN US2002137139-A1.
XX PD 26-SEP-2002.
XX PF 24-SEP-2001; 2001US-0960352.
XX PR 12-JAN-1999; 99US-115707P.
XX ER 11-JAN-2000; 2000US-0480902.
XX XX (BYAT/) BYATT J C.
XX PA (MATH/) MATHIALAGAN N.
XX PA (TAON/) TAO N.
XX PA (WAR/) WARREN W C.
XX FI Byatt JC, Mathialagan N, Tao N, Warren WC;
XX DR WPI; 2003-110599/10.
XX CC New nucleic acid associated with lactation, and muscle and fat
PT deposition, useful for genome mapping, gene identification and
PT analysis, cattle breeding, or for genetically improving cattle -
XX PS Claim 2; SEQ ID No 12478; 245pp; English.
XX CC The invention relates to a purified nucleic acid molecule associated with
CC lactation or muscle and fat deposition (designated LMFD), derived
CC from cattle, and the LMFD nucleic acid can specifically hybridise to a
CC second nucleic acid molecule comprising any of 15112 nucleotide
CC sequences, appearing as ABX4836-ABX4947, or complements of them.
CC Also included are: (1) a transformed cell having a nucleic acid
CC comprising an LMFD nucleic acid linked to a promoter and a 3' non-
CC translated sequence that functions in the cell to cause termination of
CC transcription and addition of polyadenylated ribonucleotides to a 3' end
CC of the mRNA molecule; and (2) determining a level or pattern of a
CC molecule in a bovine cell or tissue comprising: (a) incubating a marker
CC nucleic acid (comprising any of the 15112 nucleic acid sequences or its
CC complement or fragment) with a complementary nucleic acid molecule
CC obtained from the bovine cell or tissue, where hybridisation between the
CC marker nucleic acid and the complementary nucleic acid permits the
CC detection of the molecule; and (b) detecting the level or pattern of the
CC complementary nucleic acid, where the detection of the complementary
CC nucleic acid is predictive of the level or pattern of the molecule.
CC The LMFD nucleic acid is used for determining a level or pattern
CC of a molecule in a bovine cell or tissue. It is useful for genome
CC mapping, gene identification and analysis, cattle breeding, preparation
CC of constructs for use in cattle gene expression, or for genetically
CC improving cattle. The present sequence is one of the 15112 bovine
CC LMFD EST (expressed sequence tag) nucleic acids.
CC Note: The present sequence was not shown in the specification but
CC was obtained in electronic format from the USPTO web site:
CC seqdata.uspto.gov/sequence.html?DocID=20020137139.
XX SQ Sequence 434 BP; 135 A; 109 C; 86 G; 104 T; 0 other;
Query Match 92.1%; Score 74.6; DB 25; Length 434;
Best Local Similarity 95.1%; Pred. No. 2.4e-16;
Matches 77; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 GCCACATCAGATGGCTCAAGCCACACATGTCCTGCGACCCAGTGGCATGGCGCTCT 60
DB 36 GCCACATCAGATGGCTCAAGCCACACATGTCCTGCGACCCAGTGGCATGGCGCTCT 95

CC (CPSI) gene. The method is used to detect subjects at risk of
 CC hepatitis, sclerosis, pulmonary hypertension and bone marrow
 CC transplant toxicity. These conditions can be treated or prevented
 CC by administration of a nitric oxide precursor or by gene
 CC therapy (administration of sequences that encode CPSI).
 XX
 SQ Sequence 5761 BP; 1625 A; 1209 C; 1309 G; 1618 T; 0 other;

Query Match 98.0%; Score 79.4; DB 22; Length 5761;
 Best Local Similarity 98.8%; Pred. No. 8.9e-18;
 Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTGCCACCCAGTGGCGCTCT 60
 |||||
 DB 4300 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTGCCACCCAGTGGCGCTCT 4359
 |||||

QY 61 CAAGAAGGACAGATCCCGC 81
 |||||
 DB 4360 CAAGAAGGACAGATCCCGC 4380
 |||||

RESULT 12
 AAC89487
 ID AAC89487 standard; DNA; 5761 BP.
 XX
 AC AAC89487;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Human carbamyl phosphate synthase DNA.
 XX
 KW Urea cycle; carbamyl phosphate synthase I; CPSI; hepatitis;
 KW sclerosis; pulmonary hypertension; bone marrow; gene therapy; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200073322-A1.
 XX
 PD 07-DEC-2000.
 XX
 PF 01-JUN-2000; 2000WO-US15079.
 XX
 PR 01-JUN-1999; 99US-0323472.
 XX
 PA (UYVA-) UNIV VANDERBILT.
 XX
 PI Summar ML, Christman BW;
 XX
 DR WPI; 2001-049926/06.
 XX
 PT Detecting susceptibility to suboptimal urea cycle function, e.g. bone
 PT marrow transplant toxicity, comprises identifying a polymorphism in the
 PT gene for carbamyl phosphate synthase -
 XX
 PS Claim 65; Page 146; 171pp; English.
 XX
 CC The present invention relates to screening for susceptibility to
 CC suboptimal urea cycle function or to bone marrow transplant toxicity
 CC by detecting a polymorphism in the carbamyl phosphate synthase I
 CC (CPSI) gene. The method is used to detect subjects at risk of
 CC hepatitis, sclerosis, pulmonary hypertension and bone marrow
 CC transplant toxicity. These conditions can be treated or prevented
 CC by administration of a nitric oxide precursor or by gene
 CC therapy (administration of sequences that encode CPSI).
 XX
 SQ Sequence 5761 BP; 1625 A; 1209 C; 1309 G; 1618 T; 0 other;

Query Match 98.0%; Score 79.4; DB 22; Length 5761;
 Best Local Similarity 98.8%; Pred. No. 8.9e-18;
 Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTGCCACCCAGTGGCGCTCT 60
 |||||
 DB 4300 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTGCCACCCAGTGGCGCTCT 4359
 |||||

Db 4300 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTGCCACCCAGTGGCGCTCT 4359
 |||||

QY 61 CAAGAAGGACAGATCCCGC 81
 |||||
 Db 4360 CAAGAAGGACAGATCCCGC 4380
 |||||

RESULT 13
 AAC89488
 ID AAC89488 standard; DNA; 5761 BP.
 XX
 AC AAC89488;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Human carbamyl phosphate synthase DNA.
 XX
 KW Urea cycle; carbamyl phosphate synthase I; CPSI; hepatitis;
 KW sclerosis; pulmonary hypertension; bone marrow; gene therapy; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200073322-A1.
 XX
 PD 07-DEC-2000.
 XX
 PF 01-JUN-2000; 2000WO-US15079.
 XX
 PR 01-JUN-1999; 99US-0323472.
 XX
 PA (UYVA-) UNIV VANDERBILT.
 XX
 PI Summar ML, Christman BW;
 XX
 DR WPI; 2001-049926/06.
 XX
 PT Detecting susceptibility to suboptimal urea cycle function, e.g. bone
 PT marrow transplant toxicity, comprises identifying a polymorphism in the
 PT gene for carbamyl phosphate synthase -
 XX
 PS Claim 65; Page 156-161; 171pp; English.
 XX
 CC The present invention relates to screening for susceptibility to
 CC suboptimal urea cycle function or to bone marrow transplant toxicity
 CC by detecting a polymorphism in the carbamyl phosphate synthase I
 CC (CPSI) gene. The method is used to detect subjects at risk of
 CC hepatitis, sclerosis, pulmonary hypertension and bone marrow
 CC transplant toxicity. These conditions can be treated or prevented
 CC by administration of a nitric oxide precursor or by gene
 CC therapy (administration of sequences that encode CPSI).
 XX
 SQ Sequence 5761 BP; 1626 A; 1209 C; 1308 G; 1618 T; 0 other;

Query Match 98.0%; Score 79.4; DB 22; Length 5761;
 Best Local Similarity 98.8%; Pred. No. 8.9e-18;
 Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTGCCACCCAGTGGCGCTCT 60
 |||||
 Db 4300 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTGCCACCCAGTGGCGCTCT 4359
 |||||

QY 61 CAAGAAGGACAGATCCCGC 81
 |||||
 Db 4360 CAAGAAGGACAGATCCCGC 4380
 |||||

RESULT 14
 ABZ84741
 ID ABZ84741 standard; cDNA; 715 BP.
 XX
 AC ABZ84741;
 XX
 DT 14-MAY-2003 (first entry)

PF 09-MAY-1991; 91JP-0135902.
XX
PR 09-MAY-1991; 91JP-0135902.
XX

PA (HARA/) HARAGUCHI Y.
PA (MATS/) MATSUDA I.
PA (MORI/) MORI M.

XX WPI; 1993-006237/01.
DR P-PSDB; AAR30636.

XX Carbamyl phosphate synthetase I gene - used to detect diseases
PT caused by carbamyl phosphatase synthetase by using overlapped
PT cDNA contg. total human cDNA sequence

XX Claim 1; Page 6-8; 12pp; Japanese.

XX This sequence is the human carbamyl phosphate synthetase I (CPSI)
CC gene. This sequence or fragments of it may be used as probes to
CC detect mutations in the CPSI gene. CPSI deficiency disease may be
CC detected by using overlapping cDNA representing the full length cDNA
CC sequence of human CPSI. This sequence was isolated using three
CC amplified fragments derived from protein coding regions of the rat
CC CPSI cDNA as probes.

XX Sequence 5215 BP; 1484 A; 1104 C; 1205 G; 1422 T; 0 other;

Query Match 98.0%; Score 79.4; DB 14; Length 5215;
Best Local Similarity 98.8%; Pred. No. 8.7e-18;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAACGCCACCAATGTCCTGCCACCCCGATGGCGCTCT 60
DB 4295 GCCACATCAGACTGGCTCAACGCCACCAATGTCCTGCCACCCCGATGGCGCTCT 4354

QY 61 CAAGAAGGACAGAAATCCCGC 81
DB 4355 CAAGAAGGACAGAAATCCCGC 4375

RESULT 10
ABN95271
ID ABN95271 standard; DNA; 5215 BP.

AC ABN95271;

XX 13-AUG-2002 (first entry)

DE Gene #1769 used to diagnose liver cancer.

XX Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW metastatic liver tumour; cytostatic; expression profile; disease state;
KW disease progression; drug toxicity; drug efficacy; drug metabolism.

XX Homo sapiens.

XX WO200229103-A2.

XX 11-APR-2002.

XX 02-OCT-2001; 2001WO-US30589.

XX 02-OCT-2000; 2000US-237054P.

XX (GENE-) GENE LOGIC INC.

XX Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;

XX WPI; 2002-426119/45.

XX Diagnosing and detecting the progression of liver cancer,
PT hepatocellular carcinoma or metastatic liver tumor in a patient,
PT involves detecting the level of expression of two or more genes in a

PT liver tissue sample -

XX Claim 1; SEQ ID NO 1769; 298pp; English.

XX The invention relates to a novel method for diagnosing and detecting the
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC tumour in a patient, and differentiating metastatic liver cancer from
CC hepatocellular carcinoma in a patient, involving detecting the level of
CC expression of two or more genes represented in ABN93503-ABN97455 in a
CC tissue sample. The method of the invention has hepatotropic, and
CC cytostatic activity. The method is useful for diagnosing and detecting
CC the progression of liver cancer, hepatocellular carcinoma and metastatic
CC liver carcinoma in a patient. The method is useful for identifying
CC expression profiles which serve as useful diagnostic markers as well as
CC markers that can be used to monitor disease states, disease progression,
CC drug toxicity, drug efficacy and drug metabolism.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 5215 BP; 1484 A; 1102 C; 1207 G; 1422 T; 0 other;

Query Match 98.0%; Score 79.4; DB 24; Length 5215;
Best Local Similarity 98.8%; Pred. No. 8.7e-18;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAACGCCACCAATGTCCTGCCACCCCGATGGCGCTCT 60
DB 4295 GCCACATCAGACTGGCTCAACGCCACCAATGTCCTGCCACCCCGATGGCGCTCT 4354

QY 61 CAAGAAGGACAGAAATCCCGC 81

DB 4355 CAAGAAGGACAGAAATCCCGC 4375

RESULT 11
AAC89479
ID AAC89479 standard; DNA; 5761 BP.

AC AAC89479;

XX 13-MAR-2001 (first entry)

DE Human N1405 carbamyl phosphate synthase I DNA.

XX Urea cycle; carbamyl phosphate synthase I; CPSI; hepatitis;
KW scleriosis; pulmonary hypertension; bone marrow; gene therapy; ds.

XX Homo sapiens.

XX WO200073322-A1.

XX 07-DEC-2000.

XX 01-JUN-2000; 2000WO-US15079.

XX 01-JUN-1999; 99US-0323472.

XX (UYVA-) UNIV VANDERBILT.

XX Summar ML, Christman BW;

XX WPI; 2001-049926/06.

XX Detecting susceptibility to suboptimal urea cycle function, e.g. bone
PT marrow transplant toxicity, comprises identifying a polymorphism in the
PT gene for carbamylphosphate synthase -

XX Disclosure; Page 124-130; 171pp; English.

XX The present invention relates to screening for susceptibility to
CC suboptimal urea cycle function or to bone marrow transplant toxicity
CC by detecting a polymorphism in the carbamyl phosphate synthase I


```

XX KW Lung cancer; cytostatic; vaccine; gene therapy; cancer;
KW gene; ss.
XX Homo sapiens.
XX OS
XX PN US2002172952-AL.
XX PD 21-NOV-2002.
XX PF 10-JUL-2001; 2001US-0902941.
XX PR 30-JUN-1999; 99US-0346492.
XX PR 15-OCT-1999; 99US-0419356.
XX PR 17-DEC-1999; 99US-0466867.
XX PR 30-DEC-1999; 99US-0476300.
XX PR 06-MAR-2000; 2000US-0519642.
XX PR 22-MAR-2000; 2000US-0533077.
XX PR 10-APR-2000; 2000US-0546259.
XX PR 27-APR-2000; 2000US-0560406.
XX PR 05-JUN-2000; 2000US-0589194.
XX PR 11-JUL-2000; 2000US-0614124.
XX PR 29-AUG-2000; 2000US-0651563.
XX PR 08-SEP-2000; 2000US-0658824.
XX PR 26-SEP-2000; 2000US-0671325.
XX PR 06-OCT-2000; 2000US-0677419.
XX PR 30-OCT-2000; 2000US-0702705.
XX PR 13-DEC-2000; 2000US-0736457.
XX PR 03-MAY-2001; 2001US-0849626.
XX PA (CORI-) CORIXA CORP.
XX PI Henderson RA, Wang T, Watanabe Y, Johnson JC, Retter MW, Durham M;
PI Carter D, Fanger GR, Vedwick TS, Bangur CS, McNabb A;
XX WI: 2003-328427/31.
XX New polynucleotide, useful for preparing a composition for treating or
PT inhibiting development of cancer, e.g. lung cancer -
XX Example 1; SEQ ID NO 93; 82pp; English.
XX The invention describes an isolated polynucleotide comprising one of 32
CC sequences, complement or degenerate variants of them. The polynucleotide
CC is useful for preparing a composition e.g. a vaccine or for gene therapy,
CC for treating or inhibiting development of cancer, e.g. lung cancer.
CC This sequence represents a polynucleotide associated with the
CC compositions and methods for the therapy and diagnosis of lung cancer.
XX SQ Sequence 531 BP; 157 A; 139 C; 108 G; 127 T; 0 other;

Query Match 100.0%; Score 81; DB 25; Length 531;
Best Local Similarity 100.0%; Pred. No. 1.5e-18;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAAGCCCAACAATGTCCTGCCACCCAGTCGGCATGGCGGCT 60
DB 157 GCCACATCAGACTGGCTCAAGCCCAACAATGTCCTGCCACCCAGTCGGCATGGCGGCT 216
QY 61 CAAGAGGACAGAAATCCCGC 81
DB 217 CAAGAGGACAGAAATCCCGC 237

RESULT 6
AAH57467
ID AAH57467 standard; cDNA; 5195 BP.
XX AC AAH57467;
XX 10-SEP-2001 (first entry)
XX Human liver cell specific cDNA sequence SEQ ID NO:307.

```

```

XX KW Human; tissue specific; diagnosis; brain; heart; skeletal muscle;
KW lung; liver; uterus; ovary; stomach; intestine; kidney; pancreas; ss;
KW metabolic disease; developmental disease; cytostatic; immunomodulatory;
KW neuroprotective; gene therapy; cancer; immunopathology; neuropathology.
XX OS Homo sapiens.
XX PN WO200132927-A2.
XX PD 10-MAY-2001.
XX PR 02-NOV-2000; 2000WO-US30396.
XX PR 04-NOV-1999; 99US-0163508.
XX PA (INCY-) INCYTE GENOMICS INC.
XX PI Sornasse T, Seilhamer JJ, Watson GA;
XX WI: 2001-291057/30.
XX New cell and tissue specific polynucleotides useful for diagnosis,
PT prognosis or monitoring of treatments for disorders where the gene is
PT associated with a cancer, immunopathology or neuropathology -
XX Claim 1; Page 230-231; 327pp; English.
XX AAH57161 to AAH57576 represent cell and tissue specific polynucleotide
CC sequences (I). (I) can have cytostatic, immunomodulatory and
CC neuroprotective activities, and can be used in gene therapy. (I) and
CC proteins (II) encoded by then are used in high throughput screening
CC assays to select DNA molecules, RNA molecules, peptide nucleic acids,
CC mimetics, peptides, proteins, agonists, antagonists, antibodies or
CC their fragments, immunoglobulins, inhibitors, drug compounds and
CC pharmaceutical agents. Expression of (I) in a sample indicates the
CC differentiation of embryonic stem cells into a tissue selected from
CC brain, heart, kidney, liver, lung, skeletal muscle or pancreatic
CC tissues. (I) and (II) are used to produce an expression profile that
CC defines a metabolic or developmental process, treatment, condition,
CC disease or disorder. The gene profile can be used for diagnosis,
CC prognosis or monitoring of treatments and for investigating a
CC predisposition to a disorder where the gene is associated with a
CC cancer, immunopathology or neuropathology.
XX SQ Sequence 5195 BP; 1469 A; 1102 C; 1211 G; 1413 T; 0 other;

Query Match 100.0%; Score 81; DB 22; Length 5195;
Best Local Similarity 100.0%; Pred. No. 2.4e-18;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAAGCCCAACAATGTCCTGCCACCCAGTCGGCATGGCGGCT 60
DB 4310 GCCACATCAGACTGGCTCAAGCCCAACAATGTCCTGCCACCCAGTCGGCATGGCGGCT 4369
QY 61 CAAGAGGACAGAAATCCCGC 81
DB 4370 CAAGAGGACAGAAATCCCGC 4390

RESULT 7
AAC89480
ID AAC89480 standard; DNA; 5761 BP.
XX AC AAC89480;
XX 13-MAR-2001 (first entry)
XX Human T1405 carbamyl phosphate synthase I DNA.
XX Urea cycle; carbamyl phosphate synthase I; CPSI; hepatitis;
KW sclerosis; pulmonary hypertension; bone marrow; gene therapy; ds.
XX

```



```
XX WPI; 2002-164634/21.
XX Novel polynucleotide encoding a lung tumour polypeptide useful for
XX stimulating and/or expanding T cells specific for a tumour protein
XX
XX Example 1; SEQ ID No 93; 223pp; English.
XX
XX The invention describes an isolated polynucleotide and polypeptide
XX useful for stimulating and/or expanding T cells specific for a tumour
XX protein for determining the presence of a cancer in a patient. A
XX composition containing the polynucleotide and/or polypeptide is useful
XX for treating a lung cancer in a patient. The polypeptide is useful for
XX removing tumour cells from a biological sample. The polynucleotide is
XX also useful as probe or primer to detect the level of mRNA encoding a
XX tumour protein. This sequence encodes a lung tumour associated protein
XX or protein fragment, described in the method of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 531 BP; 157 A; 139 C; 108 G; 127 T; 0 other;
XX
XX Query Match 100.0%; Score 81; DB 24; Length 531;
XX Best Local Similarity 100.0%; Pred. No. 1.5e-18;
XX Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 GCCATCATGAGTGGCTCAGCGCCACCAATGTCCCTGCCACCCCATGTGCGCATGGCCGCT 60
DB 157 GCCATCATGAGTGGCTCAGCGCCACCAATGTCCCTGCCACCCCATGTGCGCATGGCCGCT 216
QY 61 CAAGAAGGACAGATCCAGC 81
DB 217 CAAGAAGGACAGATCCAGC 237
XX
RESULT 4
ACA10415
ID ACA10415 standard; cDNA; 531 BP.
AC ACA10415;
XX
XX 05-JUN-2003 (first entry)
XX Human lung cancer-associated cDNA, SEQ ID 93.
XX
XX Human; lung cancer; ss; lung tumour; cytostatic; vaccine;
XX T cell expansion; CD4; CD8.
XX
XX Homo sapiens.
XX
XX US2002197669-A1.
XX
XX 26-DEC-2002.
XX
XX 03-MAY-2001; 2001US-0849626.
XX
XX 13-DEC-2000; 2000US-0736457.
XX
XX (BANG/) BANGUR C S.
XX (FANG/) FANGER G R.
XX (WANG/) WANG A.
XX (WANG/) WANG T.
XX (SWIT/) SWITZER A P.
XX (MCNE/) MCNEILL P D.
XX (CLAP/) CLAPPER J D.
XX
XX Bangur CS, Fanger GS, Wang A, Wang T, Switzer AP, McNeill PD;
XX Clapper JD;
XX
XX WPI; 2003-352750/33.
XX
```

```
PT Novel lung cancer polynucleotide encoding lung cancer protein, useful
PT for detecting the presence of lung cancer in a patient, and in
PT pharmaceutical compositions, e.g. vaccines, for treating lung cancer
XX
XX Example 1; Page -: 72pp; English.
XX
XX The invention relates to a polynucleotide encoding a lung tumour protein,
XX comprising a sequence selected from any of the 14 sequences
XX mentioned in the specification, or a sequence (S2) mentioned in
XX specification, complement of S1, sequences that hybridise to at least 20
XX contiguous residues of S1, sequences that hybridise to S1, sequences
XX having 75% preferably 90%, identity to S1, or degenerate variants of
XX S1. Also included are an isolated polypeptide (comprising a sequence (S3)
XX selected from any one of the 4 amino acid sequences mentioned in the
XX specification, a sequence encoded by the polynucleotide, or sequences
XX having at least 70%, preferably 90%, identity to a sequence encoded by
XX the polynucleotide), an expression vector comprising the polynucleotide
XX operably linked to an expression control sequence, a host cell
XX transformed or transfected with the vector, an isolated antibody (or its
XX antigen-binding fragment) that specifically binds to the polypeptide,
XX detecting the presence of a cancer in a patient, a fusion protein
XX comprising the polypeptide, an oligonucleotide that hybridises to
XX S1 under moderately stringent conditions, stimulating and/or expanding T
XX cells specific for a tumour protein (comprising contacting T cells with
XX the polynucleotide, protein or antigen-presenting cells, under conditions
XX and for a time sufficient to permit the stimulation and/or expansion of T
XX cells) and inhibiting the development of a cancer in a patient (by
XX incubating CD4+ and/or CD8+ T cells isolated from a patient with the
XX polynucleotide, protein or antigen presenting cells that express the
XX polynucleotide, such that T cells proliferate, administering to the
XX patient an effective amount of the proliferated T cells, and thus
XX inhibiting the development of a cancer in the patient. The
XX polynucleotide, protein and cells are useful in a composition for
XX stimulating an immune response in a patient, and for treating a cancer in
XX a patient (particularly lung cancer). The oligonucleotide is useful for
XX determining the presence of a cancer in a patient. The protein and
XX oligonucleotides are useful in pharmaceutical compositions, e.g.
XX vaccines. The polynucleotide is also useful as a probe or primer for
XX nucleic acid hybridisation, and in the design and preparation of
XX ribozyme molecules for inhibiting expression of tumour polypeptides and
XX proteins in tumour cells. An amplified portion of the polynucleotide is
XX useful for isolating a full-length gene from a suitable library.
XX The present sequence is a cDNA (full length, extended or partial)
XX isolated from a library derived from lung tumour/cancer cells.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from the USPTO
XX at seqdata.uspto.gov/sequence.html?DocId=20020197669.
XX
XX Sequence 531 BP; 157 A; 139 C; 108 G; 127 T; 0 other;
XX
XX Query Match 100.0%; Score 81; DB 25; Length 531;
XX Best Local Similarity 100.0%; Pred. No. 1.5e-18;
XX Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 GCCATCATGAGTGGCTCAGCGCCACCAATGTCCCTGCCACCCCATGTGCGCATGGCCGCT 60
DB 157 GCCATCATGAGTGGCTCAGCGCCACCAATGTCCCTGCCACCCCATGTGCGCATGGCCGCT 216
QY 61 CAAGAAGGACAGATCCAGC 81
DB 217 CAAGAAGGACAGATCCAGC 237
XX
RESULT 5
ABX99366
ID ABX99366 standard; cDNA; 531 BP.
XX
XX AC ABX99366;
XX
XX 22-MAY-2003 (first entry)
XX
XX Lung cancer therapy and diagnosis associated cDNA #93.
XX
```

PT gene for carbamylphosphate synthase -
XX Disclosure; Fig 10; 171pp; English.
XX
XX The present invention relates to screening for susceptibility to
XX suboptimal urea cycle function or to bone marrow transplant toxicity
XX by detecting a polymorphism in the carbamyl phosphate synthase I
XX (CPSI) gene. The method is used to detect subjects at risk of
XX hepatitis, sclerosis, pulmonary hypertension and bone marrow
XX transplant toxicity. These conditions can be treated or prevented
XX by administration of a nitric oxide precursor or by gene
XX therapy (administration of sequences that encode CPSI).
XX
XX Sequence 495 BP; 158 A; 96 C; 85 G; 155 T; 1 other;
SQ
Query Match 100.0%; Score 81; DB 22; Length 495;
Best Local Similarity 100.0%; Pred. No. 1.4e-18;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTGCCACCCAGTGGCAGTGCCTCT 60
DB 126 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTGCCACCCAGTGGCAGTGCCTCT 185
OY 61 CAAGAAGGACAGATCCCGC 81
DB 186 CAAGAAGGACAGATCCCGC 206
RESULT 2
AAF68175
ID AAF68175 standard; cDNA; 531 BP.
AC AAF68175;
XX
XX 12-APR-2001 (first entry)
XX
XX Human lung tumour protein related nucleotide sequence SEQ ID NO:93.
XX
XX Human; lung cancer; lung tumour; lung tumour protein; gene therapy;
XX lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;
XX cytostatic; antisense inhibition; ss.
XX
XX Homo sapiens.
XX
XX WO200100828-A2.
XX
XX 04-JAN-2001.
XX
XX 30-JUN-2000; 2000WO-US18061.
XX
XX 30-JUN-1999; 99US-0346492.
XX 15-OCT-1999; 99US-0419356.
XX 17-DEC-1999; 99US-0456867.
XX 30-DEC-1999; 99US-0476300.
XX 06-MAR-2000; 2000US-0519642.
XX 22-MAR-2000; 2000US-0533077.
XX 10-APR-2000; 2000US-0546259.
XX 27-APR-2000; 2000US-0560406.
XX 05-JUN-2000; 2000US-0589184.
XX
XX (CORI-) CORIXA CORP.
XX
XX Wang T, Bangur CS, Lodes MJ, Fanger GR, Vedvick TS, Carter D;
XX Retter MW, Mannion J;
XX
XX WPI; 2001-071488/08.
XX
XX Lung tumor-associated proteins and the nucleic acids that encode them,
XX useful for preventing, diagnosing and treating lung cancer -
XX
XX Example 1; Page 181; 436pp; English.
XX
XX The present invention describes immunogenic portions of lung tumour-

CC associated proteins (I) and the nucleic acids (NAs) that encode them.
CC (I) have cytostatic activity and can be used in gene therapy, antisense
CC inhibition and in vaccines. The NAs and the lung tumour-associated
CC proteins they encode may be used in the prevention, treatment and
CC diagnosis of diseases associated with their inappropriate expression,
CC especially lung cancers. For example, the NAs may be administered to
CC treat diseases by rectifying mutations or deletions in a patient's genome
CC that affect the activity of the protein by expressing inactive proteins
CC or to supplement the patients own production of (I). Additionally, the
CC NAs may be used to produce the lung-tumour associated protein, according
CC to standard recombinant DNA methodology. Conversely, antisense NA
CC molecules may be administered to down regulate protein expression by
CC binding with the cells own genes and preventing their expression. The NA
CC and complementary sequences may also be used as DNA probes in diagnostic
CC assays to detect and quantitate the presence of similar NA sequences in
CC samples, and hence which patients may be in need of treatment for lung
CC cancer. The (I) may be used as antigens in the production of antibodies
CC and in assays to identify modulators (agonists and antagonists) of the
CC expression and activity of the protein. AAF68083 to AAF68878 and
CC AAB76848 to AAB76878 represent human lung tumour protein related
CC nucleotide and protein sequences which are used in the exemplification
CC of the present invention.
XX
XX Sequence 531 BP; 157 A; 139 C; 108 G; 127 T; 0 other;
SQ
Query Match 100.0%; Score 81; DB 22; Length 531;
Best Local Similarity 100.0%; Pred. No. 1.5e-18;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTGCCACCCAGTGGCAGTGCCTCT 60
DB 157 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTGCCACCCAGTGGCAGTGCCTCT 216
OY 61 CAAGAAGGACAGATCCCGC 81
DB 217 CAAGAAGGACAGATCCCGC 237
RESULT 3
ABK38086
ID ABK38086 standard; cDNA; 531 BP.
XX
XX ABK38086;
XX
XX 21-MAY-2002 (first entry)
XX
XX cDNA encoding clone #19065 of lung tumour protein.
XX
XX Lung tumour; cancer; T cell; immune response stimulator;
XX cytostatic; gene; ss.
XX
XX Homo sapiens.
XX
XX WO200204514-A2.
XX
XX 17-JAN-2002.
XX
XX 10-JUL-2001; 2001WO-US22058.
XX
XX 11-JUL-2000; 2000US-0614124.
XX 29-AUG-2000; 2000US-0651563.
XX 08-SEP-2000; 2000US-0658824.
XX 26-SEP-2000; 2000US-0671325.
XX 06-OCT-2000; 2000US-0677419.
XX 30-OCT-2000; 2000US-0702705.
XX 13-DEC-2000; 2000US-0736457.
XX 03-MAY-2001; 2001US-0849626.
XX
XX (CORI-) CORIXA CORP.
XX
XX Wang T, Watanabe Y, Henderson RA, Johnson JC, Retter MW;
XX Marnerakis M, Carter D, Fanger GR, Vedvick TS, Bangur CS;
XX McNabb A, Wang A, Fanger N, Switzer A, McNeill PD, Clapper JD;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: September 12, 2003, 22:19:36 ; Search time 201.5 Seconds
(without alignments)
1085.134 Million cell updates/sec

Title: US-09-585-077C-3_COPY_4300_4380
Perfect score: 81
Sequence: 1 gccacatcagctgctcaa.....aagaagacagaatccaccg 81

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues
Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :				N_Geneseq_19Jun03.*
1:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*			
2:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*			
3:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*			
4:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*			
5:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*			
6:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*			
7:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*			
8:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*			
9:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*			
10:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*			
11:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*			
12:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*			
13:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*			
14:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*			
15:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*			
16:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*			
17:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*			
18:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*			
19:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*			
20:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*			
21:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*			
22:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*			
23:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*			
24:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*			
25:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*			

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	81	100.0	495	22	Target 5' and 3' s
2	81	100.0	531	22	Human lung tumour
3	81	100.0	531	24	CDNA encoding clon
4	81	100.0	531	25	Human lung cancer-
5	81	100.0	531	25	Lung cancer therap
6	81	100.0	5195	22	Human liver cell s
7	81	100.0	5761	22	Human T1405 carbam
8	81	100.0	5772	24	Human gene express

9	79.4	98.0	5215	14	AAQ34768	hCPSI gene. Homo
10	79.4	98.0	5215	24	ABR95271	Gene #1769 used to
11	79.4	98.0	5761	22	AAC89479	Human N1405 carbam
12	79.4	98.0	5761	22	AAC89487	Human carbamyl pho
13	79.4	98.0	5761	22	AAC89488	Human carbamyl pho
14	76.8	94.8	715	25	AB84741	Toxicologically re
15	74.6	92.1	434	25	ABX47313	Bovine EST associa
16	67.4	83.2	816	24	ABA93013	Human cDNA clone B
17	64.4	79.5	459	22	AAH57259	Human liver specif
18	47.6	58.8	580	24	ABT08945	Phase-1 Rat CT gen
19	41.8	51.6	242	25	ABX36653	Bovine EST associa
20	29.4	36.3	31	22	AAI31280	Human single nucle
21	29.4	36.3	3591	22	AAI02819	Human reproductive
22	28	34.6	2153	24	ABT99890	Mouse ischaemic co
23	28	34.6	2168	18	AAV00499	Murine alpha-(2) s
24	27.4	33.8	646	24	ABT09297	Phase-1 Rat CT gen
25	27	33.3	288	25	ABX54929	Bovine EST associa
26	26.4	32.6	969	16	AAO87619	Morchella sp. DNA
27	26.4	32.6	3276	16	AAQ87604	Fungus-infected G.
28	26.4	32.6	3276	16	AAQ88050	Glucan lyase 1 gen
29	26.4	32.6	3276	20	AAV84193	Gracilariaopsis lem
30	26	32.1	1392	21	AAZ53795	Neisseria meningit
31	26	32.1	15896	21	AAA81517	N. meningitidis pa
32	26	32.1	349980	21	AAF21608	Neisseria meningit
33	26	32.1	349980	21	AAF21609	Neisseria meningit
34	26	32.1	1437668	21	AAA81490	N. meningitidis B
35	25.6	31.6	1202	19	AAV71981	A nucleotide seque
36	25.6	31.6	3345	24	ABK34183	Human CDNA differe
37	25.4	31.4	1703	21	AAZ90670	Human KCNE1 protei
38	25.4	31.4	1703	21	AAZ98902	Human long QT synd
39	25.4	31.4	1703	22	AAC89983	Human KCNE1 coding
40	25.2	31.1	284	21	AAC05908	Human secreted pro
41	25.2	31.1	8098	22	AAK90713	Human digestive sy
42	25.2	31.1	9556	22	AAK90714	Human digestive sy
43	25	30.9	25	25	ABZ84401	Toxicologically re
44	25	30.9	450	24	ABK78235	Bacillus clausili g
45	25	30.9	2372	8	AAW70814	Aspergillus niger

ALIGNMENTS

RESULT 1	
AAC89481	
ID	AAAC89481 standard; DNA; 495 BP.
AC	AAC89481;
XX	XX
DT	13-MAR-2001 (first entry)
XX	XX
DE	Target 5' and 3' sequence.
XX	XX
DE	Urea cycle; carbamyl phosphate synthase I; CPSI; hepatitis;
XX	XX
DE	sclerosis; pulmonary hypertension; bone marrow; gene therapy; ds.
XX	XX
OS	Homo sapiens.
XX	XX
PN	WO200073322-A1.
XX	XX
ED	07-DEC-2000.
XX	XX
PF	01-JUN-2000; 2000WO-US15079.
XX	XX
PR	01-JUN-1999; 99US-0323472.
XX	XX
PA	(UYVA-) UNIV VANDERBILT.
XX	XX
PI	Summar ML, Christman BW;
XX	XX
DR	WPI; 2001-049926/06.
XX	XX
PT	Detecting susceptibility to suboptimal urea cycle function, e.g. bone marrow transplant toxicity, comprises identifying a polymorphism in the

TAHVLEDTKMGYSFSGHSPSSVAGEVNTGLGGYPAITDPAYKQILITWNPILIG
NGAPDTTSLDGLSKYLSNGIKYSGLLVLDYKDYNNHMLATKSLGOWLOBEKVPA
IYGVDFRMLTKIIRKQTMGLKTEFGSQPVDFVDPNKNONLIAEVSKVYKGNPT
KYVAVCGKNNVRLYLLKQCAEVHLVPNNHDTKMEYDGLIAGGPGPALAEPLIQ
NVQKILSDKKEPLFGLSTGLNLTGLAAGAKTYKMSMANRGNQNPVINTINQOAFITA
ONCYALDNTLPAWKPLFVNVNDQTNEMHESKPPFAVOFHEVTPGPIPEYLFDF
SFTSLIKKGAFTTISVLKPAVASREVSKVLIILSGGLSIGAGEFDYSGQAVK
AMKEENVKIVLMPNNIASVOTNEVGLQADTVFLFPIPOFTEVIKAEQDGLIILGM
GGOTALNCVGLFGRVLYKGVKVLGTSVESIMATEDROLFDKLINEKINERLAPSFA
VESIEDAKAADIICYPVMIIRSAVALGGLSGICPNRETLMADISTKAFAMTQILVEK
SYTGWEIEIYEVWRADDNCVTCNMENVDANGVHTGDSVVVPAQTLNNAEQMLRR
TSYNVVRHIGIIVGECNIQIFALHPTSEMEICIEVNAKMSFNSALASKTTIGTPLAARK
IATGILPFGIKNNVSGKTSACFEPESLDIYVTKIPRWDLDRPHGTSSRIGSSMSKSGEV
MAIGRFFESFOALRMCHPSIEGTFPPLPMKNEPNSLDLRELSEPSSTRIYIAAK
AIDNMSLDEIEKLYIDKWFYKMRDILNMEKTLKGLNESMTETTKRAKEIGFSD
KOISKCLGTEROTRELRLKKNHPWKIDITLAAEYPSVTNLYVTVYNGQEHVNFDF
DHGMVILGCGPVHIGSSVEPDMCAVSSIRTLQGLKTKVTVVNCNPFVSTDEDCDKL
YFBEISLERILDIHQACGGCIISVGGQIPNNLAYPLYKNGVKIMGTSPLOIDRAED
RSIFSVDLDELKVAQPNKAVNTLINEALFEAKSDYPCLLRPSYVLSGSAMNVFSED
EMKFLTEATRVSOATPVVLFKVEGAREVEMDAVGKDRGVISHAISEHVEDAGVHSE
NATIMLPQTITISQGAIEKVKDATKAKAFASGPFNVQFLVKNVDYLVNENCLRASR
SFSVSKTGLGVDFIDVATKVLIGENVDEKHLPLDHPITIPVDYVAIKAPMFSWPLRLO
ADPTLCENASTGEVACRGEIHATPLKAMLSGFKIPQKGLIGIQOSPRPRFLGVA
BOAHNEGKLFATEATSDWLNANNVPANVPANPWSQEGQNPSSIRKLIJEDGSDILVI
NLNNNTKFDVNDYVIRTAVDSGIPLILNINFTQLFRAEAVQKSKYVDSKSLFHIRQI
SAGRAA'

BASE COUNT 1484 a 1102 c 1207 g 1422 t
ORIGIN
Query Match 98.0%; Score 79.4; DB 9; Length 5215;
Best Local Similarity 98.8%; Pred. No. 7.6e-16;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTCCACCCAGTCGCGATGCCCGTCT 60
|||||
Db 4295 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTCCACCCAGTCGCGATGCCCGTCT 4354
QY 61 CAAGAAGGACAGATCCACG 81
|||||
Db 4355 CAAGAAGGACAGATCCACG 4375

RESULT 15
AF184416
LOCUS AF184416 5761 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6346382.
ACCESSION AF184416
VERSION AF184416.1 GI:20230381
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
1 (bases 1 to 5761)
AUTHORS Summar,M.L. and Christman,B.W.
TITLE Human carbamyl phosphate synthetase I polymorphism and diagnostic
methods related thereto
JOURNAL Patent: US 6346382-A 1 12-FEB-2002;
FEATURES Location/Qualifiers
source 1..5761
/organism="unknown"
BASE COUNT 1625 a 1209 g 1618 t
ORIGIN

Query Match 98.0%; Score 79.4; DB 6; Length 5761;
Best Local Similarity 98.8%; Pred. No. 7.4e-16;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTCCACCCAGTCGCGATGCCCGTCT 60
|||||
Db 4300 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTCCACCCAGTCGCGATGCCCGTCT 4359
QY 61 CAAGAAGGACAGATCCACG 81
|||||

Db 4360 CAAGAAGGACAGATCCACG 4380

Search completed: September 12, 2003, 23:21:52
Job time : 1418.5 secs

```

QY 1 GCCATCATGACTGGCTCAAGCCCAACAATGTCCCTGCAACCCAGTGGCATGCCCGTCT 60
    |||||||
Db 30750 GCCATCATGACTGGCTCAAGCCCAACAATGTCCCTGCAACCCAGTGGCATGCCCGTCT 30691

QY 61 CAGAAGGACAGATCCCGC 81
    |||||||
Db 30690 CAGAAGGACAGATCCCGC 30670

RESULT 12
AX409122          5215 bp      DNA      linear      PAT 14-JUN-2002
LOCUS             Sequence 1769 from Patent WO0229103.
ACCESSION         AX409122
VERSION           AX409122.1 GI:21441827
KEYWORDS
SOURCE            Homo sapiens (human)
ORGANISM          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS           Alves, C., Horne, D., Peres-da-Silva, S., and Vockley, J.G.
TITLE             Gene expression profiles in liver cancer
JOURNAL           Patent: WO 0229103-A 1769 11-APR-2002;
                  GENE LOGIC INC (US)
FEATURES
  source          Location/Qualifiers
    1..5215       /organism="Homo sapiens"
                  /mol_type="genomic DNA"
                  /db_xref="taxon:9606"
                  /notes="EMBL/GenBank Accession No. D90282"
BASE COUNT       1484 a 1102 c 1207 g 1422 t
ORIGIN
Query Match      98.0%; Score 79.4; DB 6; Length 5215;
Best Local Similarity 98.8%; Pred. No. 7.6e-16;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCATCATGACTGGCTCAAGCCCAACAATGTCCCTGCAACCCAGTGGCATGCCCGTCT 60
    |||||||
Db 4295 GCCATCATGACTGGCTCAAGCCCAACAATGTCCCTGCAACCCAGTGGCATGCCCGTCT 4354

QY 61 CAGAAGGACAGATCCCGC 81
    |||||||
Db 4355 CAGAAGGACAGATCCCGC 4375

RESULT 13
E04065          5215 bp      RNA      linear      PAT 29-SEP-1997
LOCUS             CDNA encoding human carboxyl phosphate synthetase 1.
ACCESSION         E04065
VERSION           E04065.1 GI:2172275
KEYWORDS          JP 1992335889-A/1.
SOURCE            Homo sapiens (human)
ORGANISM          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS           Haraguchi, Y., Mori, M., and Matsuda, I.
TITLE             CARBOXYLPHOSPHATE SYNTHETASE I GENE, DETECTION OF MUTATION OF THE
                  GENE AND DNA PROBE TO USED THEREFOR
JOURNAL           Patent: JP 1992335889-A 1 24-NOV-1992;
                  HARAGUCHI YOGO, MORI MASATAKA, MATSUDA ICHIRO
                  OS Homo sapiens (human)
                  PN JP 1992335889-A/1
                  PD 24-NOV-1992
                  PF 09-MAY-1991 JP 1991135902
                  PI HARAGUCHI YOGO, MORI MASATAKA, MATSUDA ICHIRO PC
                  C12N15/52,C07H21/04,C12N15/11,C12Q1/68//A61B10/00; CC
                  strandedness: Double;
                  CC topology: Linear;
                  CC hypothetical: No;

```

```

CC anti-sense: No;
CC *source: tissue_type=Liver;
FH Key            Location/Qualifiers
FH 5'UTR          1..118
FH CDS            119..4621
FT 3'UTR          4622..5215.
FT CDS            1..5215
FEATURES
  source          Location/Qualifiers
    1..5215       /organism="Homo sapiens"
                  /mol_type="genomic RNA"
                  /db_xref="taxon:9606"
BASE COUNT       1484 a 1102 c 1207 g 1422 t
ORIGIN
Query Match      98.0%; Score 79.4; DB 6; Length 5215;
Best Local Similarity 98.8%; Pred. No. 7.6e-16;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCATCATGACTGGCTCAAGCCCAACAATGTCCCTGCAACCCAGTGGCATGCCCGTCT 60
    |||||||
Db 4295 GCCATCATGACTGGCTCAAGCCCAACAATGTCCCTGCAACCCAGTGGCATGCCCGTCT 4354

QY 61 CAGAAGGACAGATCCCGC 81
    |||||||
Db 4355 CAGAAGGACAGATCCCGC 4375

RESULT 14
HUMCP51          5215 bp      mRNA      linear      PRI 29-MAY-2002
LOCUS             Homo sapiens mRNA for carboxyl phosphate synthetase I, complete
DEFINITION        cds.
ACCESSION         D90282
VERSION           D90282.1 GI:219552
KEYWORDS          Homo sapiens (human)
SOURCE            Homo sapiens
ORGANISM          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE          1 (bases 1 to 5215)
AUTHORS           Haraguchi, Y., Uchino, T., Takiguchi, M., Endo, F., Mori, M. and
                  Matsuda, I.
TITLE             Cloning and sequence of a cDNA encoding human carboxyl phosphate
                  synthetase I: molecular analysis of hyperammonemia
JOURNAL           Gene 107 (2), 335-340 (1991)
MEDLINE           92084128
PUBMED            1840546
COMMENT           Submitted (18-JAN-1991) to DDBJ by:
                  Yougo Haraguchi
                  Department of Pediatrics
                  Kumamoto University Medical School
                  1-1-1 Honjo
                  Kumamoto 860
                  Japan
                  Phone: 096-344-2111 x5654
                  Fax: 096-366-3471.
FEATURES
  source          Location/Qualifiers
    1..5215       /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /db_xref="taxon:9606"
                  /tissue_type="liver"
                  /note="short arm"
                  /EC_number="6.3.4.16"
                  /codon_start=1
                  /product="carboxyl phosphate synthetase I"
                  /protein_id="BAAL4328.1"
                  /db_xref="GI:219553"
                  /translation="MRLTAFKVVRLTKTGFTNTVTHOKWFSRPGIRLLSVKQ

```

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa K., Woon P.Y., Zhao B., Frengen E., Tatenio M., Catanese J.J. and de Jong P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://pacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-430P16, 200 base pair overlap; the clone sequenced to the right is RP11-485G2, 200 base pair overlap. Actual start of this clone is at base position 195 of RP11-349G4; actual end is at base position 183502 of RP11-349G4.

RP11-349G4 contains a transposon which has been omitted from the submitted sequence. The transposon would insert after base position 170720 of this sequence.

FEATURES

source	
1. .183696	
/organism="Homo sapiens"	
/mol_type="genomic DNA"	
/db_xref="taxon:9606"	
/chromosome="2"	
/map="2"	
/clone="RP11-349G4"	
/clone_lib="RPCI-11"	
1. .244	
repeat_region	/rpt_family="L2"
repeat_region	346. .463
repeat_region	/rpt_family="MER103"
repeat_region	530. .891
repeat_region	/rpt_family="L2"
repeat_region	1835. .1933
repeat_region	2315. .2467
repeat_region	/rpt_family="Alu"
repeat_region	3061. .4062
repeat_region	/rpt_family="L2"
repeat_region	4824. .5114
repeat_region	/rpt_family="L2"
repeat_region	5380. .5846
repeat_region	/rpt_family="ERV1"
repeat_region	6941. .7483
repeat_region	/rpt_family="L1"
repeat_region	7491. .7714
repeat_region	/rpt_family="L1"
repeat_region	7704. .7951
repeat_region	/rpt_family="L1"
repeat_region	7952. .8318
repeat_region	/rpt_family="MaLR"
repeat_region	8412. .8823
repeat_region	/rpt_family="L1"
repeat_region	8858. .8961
repeat_region	/rpt_family="L1"
repeat_region	8971. .9063
repeat_region	/rpt_family="L1"
repeat_region	9566. .9812
repeat_region	/rpt_family="L1"
repeat_region	9869. .11089
repeat_region	/rpt_family="L1"
repeat_region	11288. .11763
repeat_region	/rpt_family="ERV1"
repeat_region	12296. .12734
repeat_region	/rpt_family="MaLR"
repeat_region	12749. .13082
repeat_region	/rpt_family="L1"
repeat_region	13293. .13847
repeat_region	/rpt_family="L2"
repeat_region	15165. .15620
repeat_region	/rpt_family="L2"
repeat_region	16293. .16360
repeat_region	/rpt_family="L2"
repeat_region	17445. .17990
repeat_region	/rpt_family="MIR"
repeat_region	18558. .19067
repeat_region	/rpt_family="MaLR"
repeat_region	19260. .19646
repeat_region	/rpt_family="MaLR"
repeat_region	19955. .20104
repeat_region	/rpt_family="MIR"
repeat_region	20688. .20730
repeat_region	/rpt_family="L2"
repeat_region	20744. .20983
repeat_region	/rpt_family="MIR"
repeat_region	21292. .21726
repeat_region	/rpt_family="L2"
repeat_region	23120. .23421
repeat_region	/rpt_family="Alu"
repeat_region	26196. .26367
repeat_region	/rpt_family="MIR"
repeat_region	26530. .26758
repeat_region	/rpt_family="Alu"
repeat_region	27385. .27943
misc_feature	/notes="similar to EST N58033 (NID:g1201923) yv63d03.s1"
misc_feature	27386. .28085
misc_feature	/notes="similar to EST AI936814 (NID:g5675684) wp69g09.x1"
misc_feature	27386. .28068
misc_feature	/notes="similar to EST AA676489 (NID:g2657011) ad39d05.s1"
misc_feature	27386. .27880
misc_feature	/notes="similar to EST AAL79297 (NID:g1760649) zp45b07.s1"
misc_feature	27386. .27823
misc_feature	/notes="similar to EST AV646249 (NID:g9867263) "
misc_feature	27386. .27819
misc_feature	/notes="similar to EST AAL13231 (NID:g1664935) zn66c01.s1"
misc_feature	27386. .27743
misc_feature	/notes="similar to EST AV645723 (NID:g9866737) "
misc_feature	27386. .27587
misc_feature	/notes="similar to EST AW235498 (NID:g6567887) xl13f11.x1"
misc_feature	27387. .27804
misc_feature	/notes="similar to EST AA77572 (NID:g2837051) z198b01.s1"
misc_feature	27387. .27814
misc_feature	/notes="similar to EST AAL91429 (NID:g1780108) zp83g12.s1"
misc_feature	27388. .27612
misc_feature	/notes="similar to EST N54164 (NID:g1195330) yv64b09.s1"
misc_feature	27389. .27736
misc_feature	/notes="similar to EST T52465 (NID:g654325) yb32b06.s1"
misc_feature	27391. .28148
misc_feature	/notes="similar to EST AIL32943 (NID:g6360259) "
misc_feature	27392. .27724
misc_feature	/notes="similar to EST AIL196432 (NID:g3749038) ui69f12.x1"
misc_feature	27407. .27724
misc_feature	/notes="similar to EST T69421 (NID:g680569) yc38g06.s1"
misc_feature	27486. .28021
misc_feature	/notes="similar to EST AV650985 (NID:g9871999) "
misc_feature	27487. .28199
misc_feature	/notes="similar to EST AV650994 (NID:g9872008) "
misc_feature	27488. .28121
misc_feature	/notes="similar to EST AV661872 (NID:g9882886) "
misc_feature	27513. .28021
misc_feature	/notes="similar to EST AV651545 (NID:g9872559) "
misc_feature	27519. .28148
misc_feature	/notes="similar to EST AF074674 (NID:g6470492) "
misc_feature	27582. .28199

Query Match

100.0%; Score 81; DB 9; Length 183696;

Best Local Similarity 100.0%; Pred. No. 1.3e-16;

Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
REFERENCE 1 (bases 1 to 121959)
AUTHORS Haberie,J. and Koch,H.G.
TITLE Human carbamylphosphate synthetase 1
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 121959)
AUTHORS Haberie,J. and Koch,H.G.
TITLE Direct Submission
JOURNAL Submitted (08-AUG-2002) Pediatrics, University of Munster,
Albert-Schweitzer-Str. 33, Munster 48129, Germany
FEATURES
Location/Qualifiers
1. .121959
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="2"
/map="2q35"
mRNA
join(<116..241,16680..16789,19728..19872,20803..20892,
23096..23152,25999..26091,31440..31529,33488..33616,
34182..34288,35213..35351,36261..36338,37890..37988,
38869..38964,42754..42943,43937..44094,45584..45712,
48484..48628,50113..50323,51743..51941,55499..55675,
59805..59923,81084..81225,82532..82597,83378..83441,
85866..86047,91245..91439,91855..91922,93745..93820,
97407..97484,99907..100014,101981..102070,103867..104037,
106505..106579,111568..111666,118284..118343,
119110..119222,120389..120518,121269..121367)
/product="carbamylphosphate synthetase 1"
CDS
join(<116..241,16680..16789,19728..19872,20803..20892,
23096..23152,25999..26091,31440..31529,33488..33616,
34182..34288,35213..35351,36261..36338,37890..37988,
38869..38964,42754..42943,43937..44094,45584..45712,
48484..48628,50113..50323,51743..51941,55499..55675,
59805..59923,81084..81225,82532..82597,83378..83441,
85866..86047,91245..91439,91855..91922,93745..93820,
97407..97484,99907..100014,101981..102070,103867..104037,
106505..106579,111568..111666,118284..118343,
119110..119222,120389..120518,121269..121367)
/EC_number="6.3.4.16"
/codon_start=1
/product="carbamylphosphate synthetase 1"
/protein_id="AAN7181.1"
/db_xref="GI:25992606"
/translation="MTRILTAFKVVRVLTGFTNTVAHOKWFSRPGIRLLSVKAO
TAHVLDTGKMGYSFGHPSSVAGEVNTGLGYPFAITDPAYKQILTMANPLIG
NGAPDTALDELGLSKYLENGIKVSLLDLYSKDYNHLAKSLGQWLQKTPA
IYVDRLMLKIRKQGMGLKIEFGQPFVDPKQNLRAEIVSKDVKYQKGRPT
KVAVDCGKNNVRLIKRGAEVHVPWNHDFRMEYDGLIAGGPNFALAEPLIQ
NVRKILESRRKELFLGISTGLNLTGLAAGAKYKSMANRGNQFVNITNKAFTTA
QNHGYALDNTLPAKPKPLFVNVDQTEGIMHESKPFPAVFHPVTPGIDTYEFLD
SFFSLIKKQKATITSVLPKPAVSRVSKVLIILSGGLSIGQAGEFYSGSAVK
AMKEENVKTMNPNTASVOTNEVLQADTVYFLPTPOFVTEVIAEQPDGLILGM
GGOTPALNCVELKRGVLKEYGVKVLGTSVESIMATEDQELSDKLEINEKTAQSPA
VESTEDALKAIDTIGPYMIRSAIYALGGLSGICPNRETLMDLSKAFAMTNQILVEK
SVTGKWEIEYEVRRADNDVTCVNMENVDANGVHGDSVVVYAPATLSNAEFLMR
TSINVRHPIGVEGNIQFALHPTMEYCIIEYNRLSRSLAKATGYPLAFIAAK
IALGIPPLTKNVKSVSKTSACFESLDYMTKIPRMDLDRPHGTSSRIGSMKSGVEV
MAIGRFEESFOKALRMCHPSIEGFTPLPMKWEPSNLDLRLKELSPESSTRIVATAK
AIDNNLSDEIKLTYIDKWLFLKMRDILNWEKTLKGLNSMSEETELKRAKELGSD
KQISKGLGTEAQTRELKKNINHPWKQIDTLAAEYPSVNTILVYINGQERHDFD
DHGMNVLGGPHVHSGVDFDWCVAISRTLRQKKTVVVNCNPETVSTDFDCEDKL
YFELSLELDITTYQACGGCIISVGQIPNNLAVPLFKNGVKIMGTSLQIDRAED
RSIFSAVLDELVAQAPKAVNTLNLEAFKSDVPCLLRPSVYLSGSAMNVFSED
EMKFLDEATRVQSRHVVLTKFVEGAREVEMDVGKGRVISHAISEHVEDAGVSHG
DATMLPTQISOGATEKVKDATSKIAKAFIISCPNVQFLVKGNDVILTECHLRASR
SFPFVSLIGVDEIDVAKVMIGENDEKHLPTLDHPILPADIYVAIKAPMFSWPLRD
ADPILRCMASGVACVGEHGHAFKMLSTGFKIPQKGLITIGIQSPFRPLGVA
EQLNEGFLKFAETASDNNANVPATVPANPSEQNSLRKLIRGSDILVI
NLPNNNTKFVHDNVIIRRTAVDSGITPLITNFQVTKLFAEAVOKSRKVDKSLFLHYQY
SAGKAA"
BASE COUNT 38448 a 20708 c 22753 g 40050 t
ORIGIN
Query Match 100.0%; Score 81; DB 9; Length 121959;
```

```

variation
1631
/genes="CPS1"
/notes="mutation in CPS1 deficiency"
/replace="t"
2679
/genes="CPS1"
/notes="putative polymorphism (silent)"
/replace="g"
2838
/genes="CPS1"
/notes="ambiguous"
/replace="c"
BASE COUNT 1266 a 967 c 1113 g 1157 t
ORIGIN

Query Match 100.0%; Score 81; DB 9; Length 4503;
Best Local Similarity 100.0%; Pred. No. 2.3e-16;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACATCAGCTGGCTCAACGCCAACAAATGTCCTGCCACCCAGTGGCATGGCGCTCT 60
|||||
Db 4177 GCCACATCAGCTGGCTCAACGCCAACAAATGTCCTGCCACCCAGTGGCATGGCGCTCT 4236
|||||

QY 61 CAAGAAGGACAGATCCCGCAG 81
|||||
Db 4237 CAAGAAGGACAGATCCCGCAG 4257

RESULT 8
LOCUS ARI184417 5761 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 3 from patent US 6346382.
ACCESSION ARI184417
VERSION ARI184417.1 GI:20230382
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 5761)
AUTHORS Summar,M.L. and Christman,B.W.
TITLE Human carbamyl phosphate synthetase I polymorphism and diagnostic
methods related thereto
JOURNAL Patent: US 6346382-A 3 12-FEB-2002;
FEATURES Location/Qualifiers
source 1..5761
/organism="unknown"
BASE COUNT 1624 a 1210 c 1309 g 1618 t
ORIGIN

Query Match 100.0%; Score 81; DB 6; Length 5761;
Best Local Similarity 100.0%; Pred. No. 2.2e-16;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACATCAGCTGGCTCAACGCCAACAAATGTCCTGCCACCCAGTGGCATGGCGCTCT 60
|||||
Db 4300 GCCACATCAGCTGGCTCAACGCCAACAAATGTCCTGCCACCCAGTGGCATGGCGCTCT 4359
|||||

QY 61 CAAGAAGGACAGATCCCGCAG 81
|||||
Db 4360 CAAGAAGGACAGATCCCGCAG 4380

RESULT 9
LOCUS AF154830 5761 bp mRNA linear PRI 09-JUN-1999
DEFINITION Homo sapiens carbamyl phosphate synthetase I mRNA, complete cds.
ACCESSION AF154830
VERSION AF154830.1 GI:5020419
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

variation
1631
/genes="CPS1"
/notes="mutation in CPS1 deficiency"
/replace="t"
2679
/genes="CPS1"
/notes="putative polymorphism (silent)"
/replace="g"
2838
/genes="CPS1"
/notes="ambiguous"
/replace="c"
BASE COUNT 1266 a 967 c 1113 g 1157 t
ORIGIN

Query Match 100.0%; Score 81; DB 9; Length 4503;
Best Local Similarity 100.0%; Pred. No. 2.3e-16;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACATCAGCTGGCTCAACGCCAACAAATGTCCTGCCACCCAGTGGCATGGCGCTCT 60
|||||
Db 4177 GCCACATCAGCTGGCTCAACGCCAACAAATGTCCTGCCACCCAGTGGCATGGCGCTCT 4236
|||||

QY 61 CAAGAAGGACAGATCCCGCAG 81
|||||
Db 4237 CAAGAAGGACAGATCCCGCAG 4257

RESULT 8
LOCUS ARI184417 5761 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 3 from patent US 6346382.
ACCESSION ARI184417
VERSION ARI184417.1 GI:20230382
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 5761)
AUTHORS Summar,M.L. and Christman,B.W.
TITLE Human carbamyl phosphate synthetase I polymorphism and diagnostic
methods related thereto
JOURNAL Patent: US 6346382-A 3 12-FEB-2002;
FEATURES Location/Qualifiers
source 1..5761
/organism="unknown"
BASE COUNT 1624 a 1210 c 1309 g 1618 t
ORIGIN

Query Match 100.0%; Score 81; DB 9; Length 5761;
Best Local Similarity 100.0%; Pred. No. 2.2e-16;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACATCAGCTGGCTCAACGCCAACAAATGTCCTGCCACCCAGTGGCATGGCGCTCT 60
|||||
Db 4300 GCCACATCAGCTGGCTCAACGCCAACAAATGTCCTGCCACCCAGTGGCATGGCGCTCT 4359
|||||

QY 61 CAAGAAGGACAGATCCCGCAG 81
|||||
Db 4360 CAAGAAGGACAGATCCCGCAG 4380

RESULT 10
LOCUS AF536523 121959 bp DNA linear PRI 02-DEC-2002
DEFINITION Homo sapiens carbamyl phosphate synthetase 1 gene, complete cds;
nuclear gene for mitochondrial product.
ACCESSION AF536523
VERSION AF536523.1 GI:25992605
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```


Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1
REFERENCE
AUTHORS
Wang, T., Bangur, C.S., Lodes, M.J., Fanger, G.R., Vedwick, T.S.,
Carter, D., Retter, M.W. and Mannion, J.
TITLE
Compositions and methods for the therapy and diagnosis of lung
cancer
JOURNAL
Patent: WO 0100828-A 93 04-JAN-2001;
CORIXA CORPORATION (US)
FEATURES
source
1. 531
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT
157 a 139 c 108 g 127 t
ORIGIN
Query Match 100.0%; Score 81; DB 6; Length 531;
Best Local Similarity 100.0%; Pred. No. 3.1e-16;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCCATCATGACTGGCTCAAGCCACCAATGTCCTGCCACCCAGTGGCATGGCGCTCT 60
|||||
Db 157 GCCATCATGACTGGCTCAAGCCACCAATGTCCTGCCACCCAGTGGCATGGCGCTCT 216
|||||
QY 61 CAAGAAGACAGATCCCGC 81
|||||
Db 217 CAAGAAGACAGATCCCGC 237
|||||
RESULT 6
AX367383
LOCUS
AX367383 531 bp DNA linear PAT 17-FEB-2002
DEFINITION
Sequence 93 from Patent WO204514.
ACCESSION
AX367383
VERSION
AX367383.1 GI:18855487
KEYWORDS
Homo sapiens (human)
SOURCE
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1
REFERENCE
AUTHORS
Wang, T., Watanabe, Y., Henderson, R.A., Johnson, J.C., Retter, M.W.,
Marnerakis, M., Carter, D., Fanger, G.R., Vedwick, T.S., Bangur, C.S.,
McNabb, A., Fanger, N., Switzer, A., McNeill, P.D. and Clapper, J.D.
TITLE
Compositions and methods for the therapy and diagnosis of lung
cancer
JOURNAL
Patent: WO 0204514-A 93 17-JAN-2002;
CORIXA CORPORATION (US)
FEATURES
source
1. 531
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT
157 a 139 c 108 g 127 t
ORIGIN
Query Match 100.0%; Score 81; DB 6; Length 531;
Best Local Similarity 100.0%; Pred. No. 3.1e-16;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCCATCATGACTGGCTCAAGCCACCAATGTCCTGCCACCCAGTGGCATGGCGCTCT 60
|||||
Db 157 GCCATCATGACTGGCTCAAGCCACCAATGTCCTGCCACCCAGTGGCATGGCGCTCT 216
|||||
QY 61 CAAGAAGACAGATCCCGC 81
|||||
Db 217 CAAGAAGACAGATCCCGC 237
|||||
RESULT 7
HSCFSIMUT
LOCUS
HSCFSIMUT 4503 bp mRNA linear PRI 24-SEP-1998

Homo sapiens mRNA for carbamyl phosphate synthetase 1.
X15793
ACCESSION
Y15793.1 GI:3228247
VERSION
Y15793.1
KEYWORDS
carbamoyl-phosphate synthetase I; CPS1 gene; mutation.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1
REFERENCE
AUTHORS
Finckh, U., Kohlschutter, A., Schafer, H., Sperhake, K., Colombo, J.P.
and Gal, A.
TITLE
Prenatal diagnosis of carbamoyl phosphate synthetase I deficiency
by identification of a missense mutation in CPS1
JOURNAL
Hum. Mutat. 12 (3), 206-211 (1998)
MEDLINE
98375696
PubMed
9711878
REFERENCE
2
AUTHORS
Finckh, U.
TITLE
Direct Submission
JOURNAL
Submitted (05-DEC-1997) U. Finckh, Dept. of Human Genetics, FRG
REMARK
University Hospital Eppendorf, Butenfeld 42, 22529 Hamburg, FRG
REFERENCE
3 (bases 1 to 4503)
AUTHORS
Finckh, U.
TITLE
Direct Submission
JOURNAL
Submitted (12-JUN-1998) U. Finckh, Dept. of Human Genetics, FRG
COMMENT
University Hospital Eppendorf, Butenfeld 42, 22529 Hamburg, FRG
On Jun 16, 1998 this sequence version replaced gi:2959899.
Related sequence D90282.
FEATURES
Location/Qualifiers
1. 4503
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="2"
/map="q35"
/tissue_type="liver"
1. 4503
/gene="CPS1"
1. 4503
/gene="CPS1"
/codon_start=1
/product="carbamoyl phosphate synthetase 1"
/protein_id="CA75785.1"
/db_xref="GI:3228248"
/db_xref="SWISS-PROT:P31327"
/translation="MTRILTAQVYRTLTGTGFTNTVAHQKWKFSRPGIRLLSVRAQ
TAHIVLEDTGKMGYSFGHPSVAGVEVNTGLGYPEAIDTPAYKQILTMANPIIG
NGGAPDTALDELGLSKYLESGIKVSGILLVDY SKDYNHMTATKSLGOMLEKVPV
IYGVDTMLTKIIRDKGTMLGKIEFGQPVDFDPNKNLTAETKDKVYKGNPT
KYAVDCGKNNVIRLLYKGAEVHLYPNHDETKMEYDGLIAGGPGNALAEPLIQ
NVRKILESRRKPELFGISGLNLTGLAAGATKTKMEANRQNGQVPLNITNKQAFITA
QNHGIALDNTLPAGKPLFVNVDQTNEGIMHSEKFEFFAVQFHPVTPGPTDYEYLF
SFFSLIKGKATITSVLPKALVASREYKSVLILSGGLSIGQAEFFYSQAVK
AMKENVKTVMNPNIASVQTNVEGLKQADYVFLPTPOFVETIKAEQDPDGLILGM
GGOTALNCGVELFKRGVLKXVGLGTSVESIMATEDQFLSKDLKNEKIAPSFA
VESTEDALKRAADTIGYPMIRSAVALGSLGICPNRETLMDLSTKAFATNQLIYK
SVTGWKEIYEVYRDADNCCVTCNMENVDANGVHGDSVVVPAQTLISRAEFQMLR
TSINVRHLGVEGECNIQFALHPTSMYCIIEYNARLSRSALASKATGFLAFIAK
IATGIPLEPNSVSKTSACFEPLDYMVKIPRMDLRDRHGTSTSSIGSMKSVGEV
MAIGRTPEFSQKALRMCHPTEGFTPLPMNKWPNLDRKELSPSSRTYIAIAK
AIDNNSLDEIKLTVIDKWFYLRMDILNMEKTLKGLNSFSMTTEELTKRAKIGPSD
KQISCLGLTEAQTRELKRNHPWKQIDLLAAEYSPVNYLYVYNGEHDVNF
DHGMVGGCGPHTSGSVFDFWCNASSIRTLRQKKTIVVNCNPEVTSVDEDECCKL
YFELSLEIRLDIYHQEACGGCIISVGQIPNNLAVPLYKNGVIMGTSPQIDRAED
RSIFSAVLDELAKVAQPMKAYNTLNEALEFAKSDYPCLLRPSVYLSGSAMNVFSED
EMKFELEATRYSOBHPVLAKEVEGAREVEMDVGKGRVISHAISEHVEDAGVHSG
DATMLPFTQISQGLAEKVKATRIAKAFALSPENVQFLVKGNDVLYIECNLRASR
SFPFVSKLGVDFIDVATKVMGENVDKHLPTDHPPIIPADYVAIKAPMFSPLRDR
ADPLICEASTGEVACGEGITAFKMLSTGFKIPQKGLIGIQOQSRFRPLGVA
EQLHNEGKLFATEATSDLANNVNVPATVPWPSQEQNSLSIRKLIRDGSIDLVI
NLPNNTKFFVHDNVIIRRTAVDSGIPLLTNTFQVTKLFAEAYQKSRKVDKSLSFHYRQY
SAGKAA"

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 296)
AUTHORS Funghini,S., Morrione,A. and Zammarchi,E.
TITLE Direct Submission
JOURNAL Submitted (16-OCT-2002) Dep. of Paediatrics, University of Florence, Via Luca Giordano 13, Florence 50132, Italy
FEATURES
source
1..296
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="2"
/map="2q35"
84..196
/gene="CPS1"
/number=36
BASE COUNT 93 a 48 g 91 t
ORIGIN
Query Match 100.0%; Score 81; DB 9; Length 296;
Best Local Similarity 100.0%; Pred. No. 3.4e-16;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GCCACATCAGACTGGCTCAACGCCACACATGTCCCTGCCACCCAGTGGCAGCGCTCT 60
Db 99 GCCACATCAGACTGGCTCAACGCCACACATGTCCCTGCCACCCAGTGGCAGCGCTCT 158
Qy 61 CAAGAAGGACAGAAATCCAGC 81
Db 159 CAAGAAGGACAGAAATCCAGC 179
RESULT 2
AR184418
LOCUS AR184418 495 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 5 from patent US 6346382.
ACCESSION AR184418
VERSION AR184418.1 GI:20230383
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 495)
AUTHORS Summar,M.L. and Christman,B.W.
TITLE Human carbamyl phosphate synthetase I polymorphism and diagnostic methods related thereto
JOURNAL Patent: US 6346382-A 5 12-FEB-2002;
FEATURES Location/Qualifiers
source
1..495
/organism="unknown"
BASE COUNT 158 a 96 c 85 g 155 t 1 others
ORIGIN
Query Match 100.0%; Score 81; DB 6; Length 495;
Best Local Similarity 100.0%; Pred. No. 3.1e-16;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GCCACATCAGACTGGCTCAACGCCACACATGTCCCTGCCACCCAGTGGCAGCGCTCT 60
Db 126 GCCACATCAGACTGGCTCAACGCCACACATGTCCCTGCCACCCAGTGGCAGCGCTCT 185
Qy 61 CAAGAAGGACAGAAATCCAGC 81
Db 186 CAAGAAGGACAGAAATCCAGC 206
RESULT 3
AR272381
LOCUS AR272381 531 bp DNA linear PAT 10-APR-2003
DEFINITION Sequence 93 from patent US 6504010.
ACCESSION AR272381
VERSION AR272381.1 GI:29704266

KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 531)
AUTHORS Wang,T., Bangur,C.S., Lodes,M.J., Fanger,G.R., Vedvick,T.S., Carter,D., Retter,M.W., Mannion,J. and Fan,L.
TITLE Compositions and methods for the therapy and diagnosis of lung cancer
JOURNAL Patent: US 6504010-A 93 07-JAN-2003;
FEATURES Location/Qualifiers
source
1..531
/organism="unknown"
BASE COUNT 157 a 139 c 108 g 127 t
ORIGIN
Query Match 100.0%; Score 81; DB 6; Length 531;
Best Local Similarity 100.0%; Pred. No. 3.1e-16;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GCCACATCAGACTGGCTCAACGCCACACATGTCCCTGCCACCCAGTGGCAGCGCTCT 60
Db 157 GCCACATCAGACTGGCTCAACGCCACACATGTCCCTGCCACCCAGTGGCAGCGCTCT 216
Qy 61 CAAGAAGGACAGAAATCCAGC 81
Db 217 CAAGAAGGACAGAAATCCAGC 237
RESULT 4
AR275962
LOCUS AR275962 531 bp DNA linear PAT 10-APR-2003
DEFINITION Sequence 93 from patent US 6509448.
ACCESSION AR275962
VERSION AR275962.1 GI:29709607
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 531)
AUTHORS Wang,T., Bangur,C.S., Lodes,M.J., Fanger,G.R., Vedvick,T.S., Carter,D., Retter,M.W., Mannion,J., Fan,L. and Wang,A.
TITLE Compositions and methods for the therapy and diagnosis of lung cancer
JOURNAL Patent: US 6509448-A 93 21-JAN-2003;
FEATURES Location/Qualifiers
source
1..531
/organism="unknown"
BASE COUNT 157 a 139 c 108 g 127 t
ORIGIN
Query Match 100.0%; Score 81; DB 6; Length 531;
Best Local Similarity 100.0%; Pred. No. 3.1e-16;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GCCACATCAGACTGGCTCAACGCCACACATGTCCCTGCCACCCAGTGGCAGCGCTCT 60
Db 157 GCCACATCAGACTGGCTCAACGCCACACATGTCCCTGCCACCCAGTGGCAGCGCTCT 216
Qy 61 CAAGAAGGACAGAAATCCAGC 81
Db 217 CAAGAAGGACAGAAATCCAGC 237
RESULT 5
AX062466
LOCUS AX062466 531 bp DNA linear PAT 24-JAN-2001
DEFINITION Sequence 93 from Patent WO0100828.
ACCESSION AX062466
VERSION AX062466.1 GI:12540341
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: September 12, 2003, 22:25:46 ; Search time 1416.5 Seconds
(without alignments)
2339.343 Million cell updates/sec

Title: US-09-585-077c-3_COPY_4300_4380
Perfect score: 81
Sequence: 1 gcaacatcagactggctcaa.....aagaagacagaatccacgc 81

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pla.*
- 35: em.htg.rod.*
- 36: em.htg.man.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	81	100.0	296	9	AY166970S36
2	81	100.0	495	6	AR184418
3	81	100.0	531	6	AR272381
4	81	100.0	531	6	AR275962
5	81	100.0	531	6	AX062466
6	81	100.0	531	6	AX367383
7	81	100.0	531	6	Y15793 Homo sapien
8	81	100.0	5761	6	AR184417
9	81	100.0	5761	9	AF154830 Homo sapi
10	81	100.0	121959	9	AF536523
11	81	100.0	183696	9	AC008172
12	79.4	98.0	5215	6	AX409122
13	79.4	98.0	5215	6	E04065
14	79.4	98.0	5215	9	HMCPST
15	79.4	98.0	5761	6	AR184416
16	79.4	98.0	5761	6	AR184424
17	79.4	98.0	5762	6	AR184425
18	74.6	92.1	133	10	RNCPSIX36
19	74.6	92.1	147	10	RNCPSI11
20	74.6	92.1	229369	2	AC106119
21	71.4	88.1	255581	2	AC101854
22	67.4	83.2	816	6	BD091111
23	48.6	60.0	4716	5	RCU05193
24	47.6	58.8	580	6	AX525511
25	42.6	52.6	4597	5	OMU65893
26	40	45.4	4717	5	AF006491
27	34.6	42.7	63058	2	AC133885
28	31.6	39.0	6001	5	SQUCARPSYN
29	30.6	37.8	31	6	AX249689
30	29.6	36.5	4755	5	AF119250
31	29.6	36.5	63058	2	AC133885
32	29.6	36.5	72620	2	AC121127
33	29.6	36.5	161065	9	AC124799
34	29.6	36.5	161190	9	AC005137
35	29.4	36.3	5989	5	AF169248
36	29.4	36.3	36307	9	HS36601
37	29.4	36.3	256073	9	AE006464
38	29.2	36.0	232707	2	EX233828
39	29	35.8	1101	11	G67204
40	29	35.8	141672	9	HS142F18
41	28.8	35.6	142742	9	AF121782
42	28.8	35.6	199665	9	AF064857
43	28.8	35.6	340000	9	HS21C081
44	28.6	35.3	19795	9	HS1GCMDE
45	28.6	35.3	169802	9	CNS01DY2

ALIGNMENTS

RESULT 1
AY166970S36
LOCUS
DEFINITION Homo sapiens carbanoylphosphate synthetase (CPSI) gene, exon 36;
nuclear gene for mitochondrial product.
ACCESSION AY167005.1 GI:28625483
VERSION
KEYWORDS
SEGMENT
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 296)
Punghini,S., Morrone,A. and Zammarchi,E.

[illegible]

; PRIOR APPLICATION NUMBER: 60/235,834
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/234,274
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: 60/234,223
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: 60/228,924
; PRIOR FILING DATE: 2000-08-30
; PRIOR APPLICATION NUMBER: 60/224,518
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/236,369
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/224,519
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,964
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/241,809
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/249,299
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/236,327
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/241,785
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/244,617
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 60/225,268
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/236,368
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/251,856
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/251,868
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/229,344
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/234,997
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: 60/229,343
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,345
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,287
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,513
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: 60/231,413
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/229,509
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: 60/236,367
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/237,039
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,038
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/236,370
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/236,802
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,037
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,040
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/240,960
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/239,935
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/239,937
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/241,787
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,474
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/246,532
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/249,216
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,210
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/226,681
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/225,759
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/225,213
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/227,182
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/225,214
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/235,836
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/230,438
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/215,135
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 60/225,266
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/249,218
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,208
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,213
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,212
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,207
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,245
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,244
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,217
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,211
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,215
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,264
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,214
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,297
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/232,400
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/231,242
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,081
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,080
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,414
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,244
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/233,064
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/233,063
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,397
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,399
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,401
; PRIOR FILING DATE: 2000-09-14

```

;
; FEATURE:
; OTHER INFORMATION: xaa = Unknown or Other at position 174
; OTHER INFORMATION: xaa = Unknown or Other at position 363
; OTHER INFORMATION: xaa = Unknown or Other at position 1088
; OTHER INFORMATION: xaa = Unknown or Other at position 1131
; OTHER INFORMATION: xaa = Unknown or Other at position 1217
; OTHER INFORMATION: xaa = Unknown or Other at position 1389
US-09-468-147-91
Query Match      22.1%; Score 60.5; DB 11; Length 1698;
Best Local Similarity 44.4%; Pred. No. 81;
Matches 12; Conservative 4; Mismatches 10; Indels 1; Gaps 1;

QY 3 HNEGKLFATFASDWLNANNYPANPV 29
    | | : | : | : | : | | | | |
Db 911 HRPQGHLYLTPAANFEANK-PAQFV 936

RESULT 6
US-09-764-878-129
; Sequence 129, Application US/09764878
; Patent No. US20020090615A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA121
; CURRENT APPLICATION NUMBER: US/09/764,878
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 428
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 129
; LENGTH: 70
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-878-129

Query Match      22.1%; Score 60; DB 9; Length 70;
Best Local Similarity 52.4%; Pred. No. 2.2;
Matches 11; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 31 WFSQEGQNPSLSIRKLRDG 51
    : | | | : | | | | | | | |
Db 39 YPGQHGETPSLLKIQKLRVCG 59

RESULT 7
US-09-764-860-464
; Sequence 464, Application US/09764860
; Patent No. US20020094953A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC008
; CURRENT APPLICATION NUMBER: US/09/764,860
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1198
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 464
; LENGTH: 70
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-860-464

Query Match      22.1%; Score 60; DB 9; Length 70;
Best Local Similarity 52.4%; Pred. No. 2.2;
Matches 11; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 31 WFSQEGQNPSLSIRKLRDG 51
    : | | | : | | | | | | | |
Db 39 YPGQHGETPSLLKIQKLRVCG 59

```

```

RESULT 8
US-10-079-854-129
; Sequence 129, Application US/10079854
; Publication No. US20030054368A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA121C1
; CURRENT APPLICATION NUMBER: US/10/079,854
; CURRENT FILING DATE: 2002-02-22
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 428
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 129
; LENGTH: 70
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-079-854-129

Query Match      22.1%; Score 60; DB 15; Length 70;
Best Local Similarity 52.4%; Pred. No. 2.2;
Matches 11; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 31 WFSQEGQNPSLSIRKLRDG 51
    : | | | : | | | | | | | |
Db 39 YPGQHGETPSLLKIQKLRVCG 59

RESULT 9
US-10-074-095-464
; Sequence 464, Application US/10074095
; Publication No. US2003007704A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC008C1
; CURRENT APPLICATION NUMBER: US/10/074,095
; CURRENT FILING DATE: 2002-02-14
; Prior Application Number: 09/764,860
; Prior Filing Date: 2001-01-17
; Prior Application Number: 60/179,065
; Prior Filing Date: 2000-01-31
; Prior Application Number: 60/180,628
; Prior Filing Date: 2000-02-04
; Prior Application Number: 60/214,886
; Prior Filing Date: 2000-06-28
; Prior Application Number: 60/217,487
; Prior Filing Date: 2000-07-11
; Prior Application Number: 60/225,758
; Prior Filing Date: 2000-08-14
; Prior Application Number: 60/220,963
; Prior Filing Date: 2000-07-26
; Prior Application Number: 60/217,496
; Prior Filing Date: 2000-07-11
; Prior Application Number: 60/225,447
; Prior Filing Date: 2000-08-14
; Prior Application Number: 60/218,290
; Prior Filing Date: 2000-07-14
; Prior Application Number: 60/225,757
; Prior Filing Date: 2000-08-14
; Prior Application Number: 60/226,868
; Prior Filing Date: 2000-08-22
; Prior Application Number: 60/216,647
; Prior Filing Date: 2000-07-07
; Prior Application Number: 60/225,267
; Prior Filing Date: 2000-08-14
; Prior Application Number: 60/216,880
; Prior Filing Date: 2000-07-07
; Prior Application Number: 60/225,270
; Prior Filing Date: 2000-08-14
; Prior Application Number: 60/251,869
; Prior Filing Date: 2000-12-08

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2003, 02:16:53 ; Search time 36 Seconds
(without alignments)
206.709 Million cell updates/sec

Title: US-09-585-077c-2_COPY_1380_1430
Perfect score: 272
Sequence: 1 QLNHEGFKLFAETASDWLN.....PSQEGNPSLSIKRLRDG 51

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 541936 seqs, 145912426 residues

Total number of hits satisfying chosen parameters: 541936

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PublishedApplications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES							
Result No.	Score	Query Match	Length	ID	Description		
1	67.5	24.8	344	9	US-09-815-242-5059	Sequence 5059, Ap	
2	66	24.3	1102	15	US-10-156-761-14395	Sequence 14395, A	
3	63	23.2	619	10	US-09-976-059-30	Sequence 30, Appl	
4	60.5	22.2	1228	10	US-09-117-447-2	Sequence 2, Appli	
5	60.5	22.2	1698	11	US-09-468-147-91	Sequence 91, Appl	
6	60	22.1	70	9	US-09-764-878-129	Sequence 129, Appl	
7	60	22.1	70	9	US-09-764-860-464	Sequence 464, App	
8	60	22.1	70	15	US-10-079-854-129	Sequence 129, App	
9	60	22.1	70	15	US-10-074-095-464	Sequence 464, App	
10	59.5	21.9	504	15	US-10-156-761-12417	Sequence 12417, A	
11	59	21.7	339	15	US-10-169-048-40	Sequence 40, Appl	
12	58	21.3	1018	11	US-09-494-359-3	Sequence 3, Appli	
13	58	21.3	1113	9	US-09-836-470B-3	Sequence 3, Appli	
14	58	21.3	1113	15	US-10-284-138-3	Sequence 3, Appli	
15	58	21.3	1113	15	US-10-284-334-3	Sequence 3, Appli	

16	57.5	21.1	1435	15	US-10-128-714-8125	Sequence 8125, Ap
17	57	21.0	100	10	US-09-872-523-8	Sequence 8, Appli
18	56.5	20.8	576	14	US-10-029-180-102	Sequence 102, App
19	56	20.6	1113	10	US-09-738-626-5279	Sequence 5279, Ap
20	55.5	20.4	162	15	US-10-156-761-13789	Sequence 13789, A
21	55.5	20.4	232	15	US-10-156-761-11422	Sequence 11422, A
22	55.5	20.4	823	15	US-10-177-293-240	Sequence 240, App
23	55	20.2	1636	12	US-10-202-167-2	Sequence 2, Appli
24	54.5	20.0	639	10	US-09-906-209-2	Sequence 2, Appli
25	54.5	20.0	673	15	US-10-157-031-291	Sequence 291, App
26	54	19.9	441	9	US-09-879-957-34	Sequence 34, Appli
27	54	19.9	570	9	US-09-815-242-10760	Sequence 10760, A
28	54	19.9	1012	10	US-09-285-385C-4	Sequence 4, Appli
29	53.5	19.7	766	15	US-10-156-761-10075	Sequence 10075, A
30	53.5	19.7	1708	11	US-09-468-147-166	Sequence 166, App
31	53	19.5	140	12	US-10-238-075-984	Sequence 984, App
32	53	19.5	274	12	US-09-882-227-196	Sequence 196, App
33	53	19.5	395	15	US-10-128-714-3592	Sequence 3592, Ap
34	53	19.5	620	15	US-10-128-714-8592	Sequence 8592, Ap
35	52.5	19.3	311	15	US-10-156-761-14936	Sequence 14936, A
36	52.5	19.3	346	15	US-10-102-806-577	Sequence 577, App
37	52	19.1	644	10	US-09-943-692-2	Sequence 2, Appli
38	52	19.1	644	15	US-10-222-441-2	Sequence 2, Appli
39	52	19.1	644	15	US-10-229-346-2	Sequence 2, Appli
40	52	19.1	844	14	US-10-055-364-40	Sequence 40, Appl
41	52	19.1	1601	10	US-09-882-027-40	Sequence 40, Appl
42	52	19.1	3739	10	US-09-861-289-33	Sequence 33, Appl
43	52	19.1	3739	10	US-09-860-846-33	Sequence 33, Appl
44	52	19.1	3739	11	US-09-988-384B-33	Sequence 33, Appl
45	52	19.1	3739	11	US-09-836-821-33	Sequence 33, Appl

ALIGNMENTS

RESULT 1
US-09-815-242-5059
; Sequence 5059, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5059
; LENGTH: 344
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa

```
; SEQ ID NO 4
; LENGTH: 1012
; TYPE: PRT
; ORGANISM: mouse
US-09-285-385C-4

Query Match      19.9%; Score 54; DB 4; Length 1012;
Best Local Similarity 23.7%; Pred.No. 1.5e+02;
Matches 14; Conservative 7; Mismatches 26; Indels 12; Gaps 2;

QY      2 LHNEGFKLFAEATSDWL-----NANNVPANVPANPSQGNPSLSIRKLIIRDG 51
      | : ||| : || : : || : || : || : || : || : ||
Db      65 LDEDDLKLFHDRAEDWTXPKDKPGHDTGGLETSARWPDNTASNASIQAPRK---DG 120

RESULT 15
US-09-462-606-2
; Sequence 2, Application US/09462606
; Patent No. 6432408
; GENERAL INFORMATION:
; APPLICANT: MENG, XIANG-JIN
; APPLICANT: Emerson, Suzanne U.
; APPLICANT: Purcell, Robert H.
; TITLE OF INVENTION: A SWINE HEPATITIS E VIRUS AND USES THEREOF
; FILE REFERENCE: 20264267US1
; CURRENT APPLICATION NUMBER: US/09/462,606
; PRIOR FILING DATE: 2000-06-12
; PRIOR APPLICATION NUMBER: US 60/053069
; PRIOR FILING DATE: 1997-07-18
; PRIOR APPLICATION NUMBER: PCT/US98/14665
; PRIOR FILING DATE: 1998-07-17
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1708
; TYPE: PRT
; ORGANISM: Hepatitis E virus
US-09-462-606-2

Query Match      19.7%; Score 53.5; DB 4; Length 1708;
Best Local Similarity 42.3%; Pred.No. 3.4e+02;
Matches 11; Conservative 3; Mismatches 11; Indels 1; Gaps 1;

QY      3 HNEGFKLFAEATSDMLNANNVPANP 28
      | :| : || : || : || : ||
Db      921 HPGDELVLTEPAAWFEANK-PAQP 945

Search completed: September 13, 2003, 02:17:33
Job time : 22 secs
```

```
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-403-618A-8
Query Match          20.0%; Score 54.5; DB 4; Length 878;
Best Local Similarity 29.8%; Pred No. 1.1e+02;
Matches 14; Conservative 6; Mismatches 20; Indels 7; Gaps 2;

QY 11 ATEATSDWL-NANNVPANVPQSGQNPQLSSIRKL-----IRD 50
Db 210 SANSTRWLVLDINNETKLPLANDGLAEBKPRLDSDISIVYELHIRD 256

RESULT 12
US-08-630-915A-34
; Sequence 34, Application US/08630915A
; Patent No. 6309820
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; APPLICANT: HOFFMAN, No. 6309820h
; APPLICANT: KAY, Brian K.
; APPLICANT: FOWLKES, Dana M.
; APPLICANT: MCCONNELL, Stephen J.
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
; TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
; TITLE OF INVENTION: USING SAME
; NUMBER OF SEQUENCES: 227
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,915A
; FILING DATE: 03-APR-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-174
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 441 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-630-915A-34
Query Match          19.9%; Score 54; DB 4; Length 441;
Best Local Similarity 37.5%; Pred No. 53;
Matches 9; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 18 WLNANNVPANVPQSGQNPQL 41
Db 283 WFRSTGGMNKPQFEWNPD 306

RESULT 13
US-09-107-532A-6689
; Sequence 6689, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; APPLICANT: HOFFMAN, No. 6309820h
; APPLICANT: KAY, Brian K.
; APPLICANT: FOWLKES, Dana M.
; APPLICANT: MCCONNELL, Stephen J.
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
; TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
; TITLE OF INVENTION: USING SAME
; NUMBER OF SEQUENCES: 227
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneka
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 6689:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 569 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...569
; SEQUENCE DESCRIPTION: SEQ ID NO: 6689:
US-09-107-532A-6689
Query Match          19.9%; Score 54; DB 4; Length 569;
Best Local Similarity 23.9%; Pred No. 73;
Matches 11; Conservative 12; Mismatches 23; Indels 0; Gaps 0;

QY 2 LHNEGKLFATEATSDWLNANNVPANVPQSGQNPQLSSIRKL 47
Db 466 LQNDQLTMKAKTQDYWLHAKNIPGSHVLIKSDKPSDETITEAAEL 511

RESULT 14
US-09-285-385C-4
; Sequence 4, Application US/09285385C
; Patent No. 6579702
; GENERAL INFORMATION:
; APPLICANT: Greenspan, Daniel S.
; APPLICANT: Scott, Ian C.
; APPLICANT: Thomas, Christina L.
; TITLE OF INVENTION: MAMMALIAN TOLLOID-LIKE GENE AND PROTEIN
; FILE REFERENCE: 960296.96111
; CURRENT APPLICATION NUMBER: US/09/285,385C
; CURRENT FILING DATE: 1999-04-02
; PRIOR APPLICATION NUMBER: 60/111873
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 60/080550
; PRIOR FILING DATE: 1998-04-03
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
```

```
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5495

Query Match      23.2%; Score 63; DB 4; Length 228;
Best Local Similarity 26.0%; Pred. No. 1.5;
Matches 13; Conservative 11; Mismatches 22; Indels 4; Gaps 1;

QY 1 QLNHEGKFLFATSDMLNANNVPANVPVWPSQEGQNPSSLIRKLRD 50
   :| | | : | :| | :| :| :| :| :| :| :| :| :| :|
Db 157 LIKEDFDILGIEVINWLTGKNIKSKDIS----EGFNVEIKRIELLKD 202

RESULT 8
US-09-252-991A-20252
; Sequence 2052, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 20252
; LENGTH: 167
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20252

Query Match      22.4%; Score 61; DB 4; Length 167;
Best Local Similarity 37.9%; Pred. No. 1.9;
Matches 11; Conservative 4; Mismatches 14; Indels 0; Gaps 0;

QY 12 TEATSDMLNANNVPANVPVWPSQEGQNPFS 40
   :| | :| :| :| | | | :| |
Db 127 SSATAPWSPRPSAAPANSAGWPSSRRWPS 155

RESULT 9
US-09-252-991A-27861
; Sequence 27861, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27861
; LENGTH: 96
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27861

Query Match      21.3%; Score 58; DB 4; Length 96;
Best Local Similarity 40.6%; Pred. No. 2.4;
Matches 13; Conservative 3; Mismatches 16; Indels 0; Gaps 0;

QY 14 ATSDMLNANNVPANVPVWPSQEGQNPSSLIR 45
   :| | :| :| | | | | | | |
Db 47 SSSPWFSAIRWPATSWAITSFACTRSCSKRY 78

RESULT 10
US-09-629-616-3
; Sequence 3, Application US/09629616
; Patent No. 6255086
; GENERAL INFORMATION:
; APPLICANT: KUNABARA, Yoko
; APPLICANT: HASHIGUCHI, Kenichi
; APPLICANT: NAKAMATSU, Tsuyoshi
; APPLICANT: KURAHASHI, Osamu
; APPLICANT: MORI, Yukiko
; APPLICANT: ITO, Hisao
; TITLE OF INVENTION: CARBAMOYL-PHOSPHATE SYNTHETASE GENE OF CORYNEFORM
; FILE REFERENCE: OP945CIP
; CURRENT APPLICATION NUMBER: US/09/629,616
; CURRENT FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1113
; TYPE: PRT
; ORGANISM: Brevibacterium lactofermentum
US-09-629-616-3

Query Match      21.3%; Score 58; DB 3; Length 1113;
Best Local Similarity 31.4%; Pred. No. 50;
Matches 16; Conservative 8; Mismatches 27; Indels 0; Gaps 0;

QY 1 QLNHEGKFLFATSDMLNANNVPANVPVWPSQEGQNPSSLIRKLRD 51
   :| | :| | | | | :| :| :| :| :| :| :| :| :| :|
Db 994 RLASMGYKILATEGTAGMLRRNGIDCEVVLKASDIRGVEGKSIDVRIREG 1044

RESULT 11
US-09-403-618A-8
; Sequence 8, Application US/09403618A
; Patent No. 6429358
; GENERAL INFORMATION:
; APPLICANT: BROGLIE, KAREN E.
; TITLE OF INVENTION: CORN PULLULANASE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: U.S.A.
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 95
; SOFTWARE: MICROSOFT WORD FOR WINDOWS 95 (7.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/403,618A
; FILING DATE: 05-NO. 6429358-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/045,723
; FILING DATE: MAY 6, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: MAJARIAN, WILLIAM R.
; REGISTRATION NUMBER: 41,173
; REFERENCE/DOCKET NUMBER: BB-1108
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 878 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
```

```

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074.788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094.190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27544
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27544

Query Match          24.8%; Score 67.5; DB 4; Length 349;
Best Local Similarity 37.1%; Pred.No.0.67;
Matches 13; Conservative 7; Mismatches 14; Indels 1; Gaps 1;

QY      1 QLNHEGFKLFATEATSDML-NANNVPANPVAMPSQ 34
        :L::||| | | ||| : |::|||::
Db       56 RVHSEGEVFLQNAWQDWLATGRSHDMLPIAFTR 100

RESULT 6
US-09-134-001C-4694
; Sequence 4694, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134.001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064.964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055.779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4694
; LENGTH: 228
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4694

Query Match          23.2%; Score 63; DB 4; Length 228;
Best Local Similarity 26.0%; Pred. No. 1.5;
Matches 13; Conservative 11; Mismatches 22; Indels 4; Gaps 1;

QY      1 QLNHEGFKLFATEATSDMLNANNVPANPVAMPSQGQNPSSLSKRLIRD 50
        :L | | : | | | : | : | : | | : | | : | | : | |
Db       157 KLRDEDFILGIENVINWLTGSNIKSKDIS---EGTNVEIKELLLKD 202

RESULT 7
US-09-134-001C-5495
; Sequence 5495, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134.001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064.964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055.779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5495
; LENGTH: 228

```



```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96051398; PubMed=7584044;
RA Nomura N., Nagase T., Miyajima N., Sazuka T., Tanaka A., Sato S.,
RA Seki N., Kawabayashi Y., Ishikawa K., Tabata S.;
RT "Prediction of the coding sequences of unidentified human genes. II.
RT The coding sequences of 40 new genes (K1AA0041-K1AA0080) deduced by
RT analysis of cDNA clones from human cell line KG-1.";
RL DNA Res. 1:223-229(1994).
DR EMBL; D38521; BAA07526.1; -
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 1798 AA; 206406 MW; 6FCB2F2AA77999C1 CRC64;

Query Match 23.3%; Score 63.5; DB 4; Length 1798;
Best Local Similarity 37.8%; Pred. No. 66;
Matches 14; Conservative 7; Mismatches 9; Indels 7; Gaps 2;

QY 1 QLNHEGFKLPATEATSDWLNANNVPANVPWPVSEGG 37
:|||||: | |:: | | |
Db 180 ELHHKGFKLWFDLIGLWVSVQNLP-----QW---EGQ 209

RESULT 15
Q9ESY1 PRELIMINARY; PRT; 1803 AA.
AC Q9ESY1;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE TEMO.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Mruk D., Mo M.-Y., Cheng C.Y.;
RT "TEMO is a marker to study sertoli-germ cell interactions: cloning and
RT regulation of a novel testicular molecule.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF296169; AAG09060.1; -
SQ SEQUENCE 1803 AA; 206368 MW; ED73B36E135F25B2 CRC64;

Query Match 23.3%; Score 63.5; DB 11; Length 1803;
Best Local Similarity 37.8%; Pred. No. 66;
Matches 14; Conservative 7; Mismatches 9; Indels 7; Gaps 2;

QY 1 QLNHEGFKLPATEATSDWLNANNVPANVPWPVSEGG 37
:|||||: | |:: | | |
Db 185 ELHHKGFKLWFDLIGLWVSVQNLP-----QW---EGQ 214

```

Search completed: September 13, 2003, 02:15:42
Job time : 65.5 secs

[illegible]


```

ID O93337 PRELIMINARY; PRT; 2275 AA.
AC O93337;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DE Probable acetylpolysaccharide synthase
DE PyABCN (EC 6.3.5.5).
GN PYABCN
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=162425;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FGSC 4;
RA Aleksenko A., Liu W., Gojkovic Z., Nielsen J., Piskur J.;
RT "Structural and transcriptional analysis of the pyABCN, pyrd and pyrf
RT genes in Aspergillus nidulans and the evolutionary origin of fungal
RT dihydroorotases."
RL Submitted (DRC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF112473; AAD09129.1; -.
DR HSP; P00968; ICS0.
DR InterPro; IPR006220; Anth_synthII.
DR InterPro; IPR006130; Asp/Orrn_Cotranf.
DR InterPro; IPR002082; Asp_carb_mitransf.
DR InterPro; IPR006275; CarA_L_glu.
DR InterPro; IPR006274; CarA_small.
DR InterPro; IPR005483; CPase_L.
DR InterPro; IPR005479; CPase_L_D2.
DR InterPro; IPR005480; CPase_L_D3.
DR InterPro; IPR005481; CPase_L_N.
DR InterPro; IPR002474; CPase_sm_chain.
DR InterPro; IPR001317; CPS_GATase.
DR InterPro; IPR000991; GATase_L.
DR InterPro; IPR004362; MGS_like.
DR InterPro; IPR006131; OTCace_O.
DR InterPro; IPR006132; OTCace_P.
DR Pfam; PF00289; CPase_L_chain; 2.
DR Pfam; PF02786; CPase_L_D2; 3.
DR Pfam; PF02787; CPase_L_D3; 1.
DR Pfam; PF00988; CPase_sm_chain; 1.
DR Pfam; PF00117; GATase; 1.
DR Pfam; PF02142; MGS; 1.
DR Pfam; PF00185; OTCace; 1.
DR Pfam; PF02729; OTCace_N; 1.
DR PRINTS; PR00097; ANTISYTHASEII.
DR PRINTS; PR00300; ROTCASE.
DR PRINTS; PR00098; GATASE.
DR PRINTS; PR00099; CPSEGATASE.
DR PRINTS; PR00096; GATASE.
DR TIGRFAMS; TIGR00670; asp_carb_tr; 1.
DR TIGRFAMS; TIGR01368; CPaseII_small; 1.
DR TIGRFAMS; TIGR01369; CPaseII_lrg; 1.
DR PROSITE; PS00097; CARBAMOYLTRANSFERASE; 1.
DR PROSITE; PS00866; CPASE_1; 2.
DR PROSITE; PS00867; CPASE_2; 2.
DR PROSITE; PS00442; GATASE_TYPE_1; 1.
KW Ligase.
SQ SEQUENCE 2275 AA; 249739 MW; 27BA9C1FA751436A CRC64;

Query Match 25.4%; Score 69; DB 3; Length 2275;
Best Local Similarity 52.0%; Pred. No. 17;
Matches 13; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 QLHNEGKFLFATATSDWLNANNVP 25
: : : : : : : : : : : : : : : :
Db 1422 KLRDVGFLFATSGTADFLKGV 1446

RESULT 9
Q916H0 PRELIMINARY; PRT; 344 AA.
ID Q916H0
AC Q916H0;
DT 01-MAR-2001 (TReMBLrel. 16, Created)

```

```

DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Probable acetylpolysaccharide synthase
GN PA0321.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Hickey M.J., Brinkman F.S.L., Hurnagie W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goutry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Jarbig K., Lam R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
DR EMBL; AE004470; AAG03710.1; -.
DR InterPro; IPR000286; His_deacetylase.
DR Pfam; PF00850; Hist_deacetyl; 1.
DR PRINTS; PR01270; HDASUPER.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 344 AA; 37883 MW; 8238194D8911DOCE CRC64;

Query Match 24.8%; Score 67.5; DB 16; Length 344;
Best Local Similarity 37.1%; Pred. No. 3.1;
Matches 13; Conservative 7; Mismatches 14; Indels 1; Gaps 1;

QY 1 QLHNEGKFLFATATSDWLNANNVPANPVPWPSQ 34
: : : : : : : : : : : : : : : :
Db 61 RVHSEGFVRELFQAWQDWLGTGRSHDMLPIAWPTR 95

RESULT 10
Q8UG58 PRELIMINARY; PRT; 863 AA.
ID Q8UG58
AC Q8UG58;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE DNA-directed RNA polymerase.
GN ATU1183 OR AGP_C_2186.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Glendinning J., Deatherage G., Gillet W., Grant C.,
RA Kutayin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58."
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmlel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,

```

```
QY 1 QLNHEGFKLFATSDWLNANNVPANVPWPSQEQGNPSLSIRKLIRDG 51
||| :||:||||| ||| ||| :||: ||| :||| |||
Db 1381 QLNHEGFKLYATGISTWLNVDVPTAPVSWPTAEDHSSAPSTFKLIHDG 1431

RESULT 6
Q9PTW9 PRELIMINARY; PRT; 1504 AA.
AC Q9PTW9;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Acetylglutamate-activated carbamoyl phosphate synthase III.
OS Opsanus beta.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Paracanthopterygii; Batrachoididae; Opsanus.
OX NCBI_TaxID=93145;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Kong H., Kahatapitiya N., Kingsley K., Salo W.L., Anderson P.M.,
RA Wang Y.S., Walsh P.J.;
RT "Induction of expression of carbamoyl phosphate synthetase III and
RT glutamine synthetase mRNA during confinement stress in gulf toadfish
RT (Opsanus beta).";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF169248; AAD51318.2; -.
DR HSSP; P00968; 1A9X.
DR InterPro; IPR006275; CarA_L.glu.
DR InterPro; IPR006274; CarA_small.
DR InterPro; IPR005483; CPase_L.
DR InterPro; IPR005479; CPase_L_D2.
DR InterPro; IPR005480; CPase_L_D3.
DR InterPro; IPR005481; CPase_L_N.
DR InterPro; IPR005481; CPase_L_N.
DR InterPro; IPR002474; CPase_sm_chain.
DR InterPro; IPR001317; CPS_GATase.
DR InterPro; IPR000991; GATase_L.
DR InterPro; IPR004362; MGS_like.
DR Pfam; PF02786; CPase_L_chain; 2.
DR Pfam; PF02787; CPase_L_D2; 2.
DR Pfam; PF02787; CPase_L_D3; 1.
DR Pfam; PF00988; CPase_sm_chain; 1.
DR Pfam; PF00117; GATase; 1.
DR Pfam; PF02142; MGS; 1.
DR PRINTS; PR00098; CPSASE.
DR PRINTS; PR00099; CPSGATASE.
DR TIGRFAMs; TIGR01368; CPSaseIismall; 1.
DR PROSITE; PS00866; CPSASE_1; 2.
DR PROSITE; PS00867; CPSASE_2; 2.
DR PROSITE; PS00867; CPSASE_2; 2.
DR PROSITE; PS00482; DIHYDROROTASE_1; 1.
DR PROSITE; PS00483; DIHYDROROTASE_2; 1.
DR PROSITE; PS00442; GATASE_TYPE_I; 1.
SQ SEQUENCE 1504 AA; 165446 MW; C46D618F65AA3F8D CRC64;

Query Match 57.5%; Score 156.5; DB 13; Length 1504;
Best Local Similarity 60.8%; Pred. No. 9.9e-11;
Matches 31; Conservative 10; Mismatches 9; Indels 1; Gaps 1;

QY 1 QLNHEGFKLFATSDWLNANNVPANVPWPSQEQGNPSLSIRKLIRDG 51
||| :||:||||| ||| ||| :||: ||| :||| |||
Db 1380 QLNHEGFKLYATGISTWLNVDVPTAPVSWPTAEDHSSAPSTFKLIHDG 1429

RESULT 7
Q18990 PRELIMINARY; PRT; 2198 AA.
AC Q18990;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE D2085.1 protein.
GN D2085.1.

OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Baynes C.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z54284; CAA91059.1; -.
DR HSSP; P00479; 3CSU.
DR WormPep; D2085.1; CE03105.
DR InterPro; IPR006680; Amidohydro_L.
DR InterPro; IPR006220; Anth_synthII.
DR InterPro; IPR006130; Asp/Orn_Cotransf.
DR InterPro; IPR002082; Asp_carbamyltransf.
DR InterPro; IPR006275; CarA_L.glu.
DR InterPro; IPR006274; CarA_small.
DR InterPro; IPR005483; CPase_L.
DR InterPro; IPR005479; CPase_L_D2.
DR InterPro; IPR005480; CPase_L_D3.
DR InterPro; IPR005481; CPase_L_N.
DR InterPro; IPR002474; CPase_sm_chain.
DR InterPro; IPR001317; CPS_GATase.
DR InterPro; IPR002195; Dihydroorotase.
DR InterPro; IPR000991; GATase_L.
DR InterPro; IPR004362; MGS_like.
DR InterPro; IPR006131; OTCace_O.
DR InterPro; IPR006132; OTCace_P.
DR InterPro; IPR005847; Urease.
DR Pfam; PF01979; Amidohydro_L; 1.
DR Pfam; PF00289; CPase_L_chain; 2.
DR Pfam; PF02786; CPase_L_D2; 2.
DR Pfam; PF02787; CPase_L_D3; 1.
DR Pfam; PF00988; CPase_sm_chain; 1.
DR Pfam; PF00117; GATase; 1.
DR Pfam; PF02142; MGS; 1.
DR Pfam; PF00185; OTCace; 1.
DR Pfam; PF02723; OTCace_N; 1.
DR PRINTS; PR00097; AMTSNTHASEII.
DR PRINTS; PR00100; NOTCASE.
DR PRINTS; PR00098; CPSASE.
DR PRINTS; PR00099; CPSGATASE.
DR PRINTS; PR00096; GATASE.
DR ProDom; PD000518; Urease; 1.
DR TIGRFAMs; TIGR00670; asp_carb_tr; 1.
DR TIGRFAMs; TIGR01368; CPSaseIismall; 1.
DR TIGRFAMs; TIGR01369; CPSaseI_lrg; 1.
DR PROSITE; PS00097; CARBAMOYLTRANSFERASE; 1.
DR PROSITE; PS00866; CPSASE_1; 2.
DR PROSITE; PS00867; CPSASE_2; 2.
DR PROSITE; PS00482; DIHYDROROTASE_1; 1.
DR PROSITE; PS00483; DIHYDROROTASE_2; 1.
DR PROSITE; PS00442; GATASE_TYPE_I; 1.
SQ SEQUENCE 2198 AA; 242567 MW; 2C029AFD34C71A4E CRC64;

Query Match 29.4%; Score 80; DB 5; Length 2198;
Best Local Similarity 33.3%; Pred. No. 0.69;
Matches 11; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

QY 6 GFKLFATEATSDWLNANNVPANVPWPSQEQGN 38
||| :||:||||| ||| ||| :||: ||| :||| |||
Db 1349 GVELYGSKGATDYFQSNKINVKPVDWPFEGSS 1381

RESULT 8
O93937
```

```

DR InterPro; IPR000991; GATase.1.
DR InterPro; IPR004362; MGS_like.
DR Pfam; PF00289; CPase_L_chain; 2.
DR Pfam; PF02786; CPase_L_D2; 2.
DR Pfam; PF02787; CPase_L_D3; 1.
DR Pfam; PF00988; CPase_sm_chain; 1.
DR Pfam; PF00117; GATase; 1.
DR Pfam; PF02142; MGS; 1.
DR PRINTS; PR00099; CPASASE.
DR TIGRFAMS; TIGR01368; CPaseIIsmall; 1.
DR TIGRFAMS; TIGR01369; CPaseII_lrg; 1.
DR PROSITE; PS00866; CPASE_1; 2.
DR PROSITE; PS00867; CPASE_2; 2.
SQ SEQUENCE 1505 AA; 164592 MW; 89C4A22285166CA2 CRC64;

Query Match      60.3%; Score 164; DB 13; Length 1505;
Best Local Similarity 64.7%; Pred. No. 1.1e-11;
Matches 33; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

QY 1 QLNHEGKLPATDTSWLNANNVPANVPWPSQGNPNSLSIRKLRDG 51
   || |||||:||||| |||||:||||| |||||: || |||||: ||
Db 1380 QLEEGEKLYATEATSAWLCANDVPAIPVAPVTGMGDNSLPSIKRLISEG 1430
   (GD-CPase III).

RESULT 5
Q92115 PRELIMINARY; PRT; 1502 AA.
ID Q92115; Q91438;
AC Q92115; 091438; (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Carbamoyl phosphate synthetase III [glutamine] mitochondrial precursor
DE (EC 6.3.5.5) (Glutamine-dependent carbamoyl-phosphate synthetase III)
DE (GD-CPase III).
GN CPSIII.
OS Squalus acanthias (Spiny dogfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Squala; Squaloidae; Squalidae; Squalus.
OX NCBI_TaxID=7797;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 39-47.
RC TISSUE=LIVER;
RX MEDLINE=95018247; PubMed=7932737;
RA Hong J., Salo W.L., Lusty C.J., Anderson P.M.;
RT "Carbamyl phosphate synthetase III, an evolutionary intermediate in
RT the transition between glutamine-dependent and ammonia-dependent
RT carbamyl phosphate synthetases.";
RL J. Mol. Biol. 243:131-140(1994).
RN [2]
RP SEQUENCE OF 1-42 FROM N.A.
RC TISSUE=TESTIS;
RX MEDLINE=97149029; PubMed=8995057;
RA Hong J., Salo W.L., Chen Y., Atkinson B.G., Anderson P.M.;
RT "The promoter region of the carbamoyl-phosphate synthetase III gene of
RT Squalus acanthias.";
RL J. Mol. Evol. 43:602-609(1996).
RN [3]
RP ENZYME ACTIVITY, AND REGULATION.
RC TISSUE=LIVER;
RX MEDLINE=80168997; PubMed=6245445;
RA Anderson P.M.;
RT "glutamine- and N-acetylglutamate-dependent carbamoyl phosphate
RT synthetase in elasmobranchs.";
RL Science 208:291-293(1980).
RN [4]
RP ENZYME ACTIVITY, AND SUBUNITS.
RC TISSUE=LIVER;
RX MEDLINE=82053090; PubMed=7298655;
RA Anderson P.M.;
RT "Purification and properties of the glutamine- and N-acetyl-L-
RT glutamate-dependent carbamoyl phosphate synthetase from liver of
RT Squalus acanthias.";
```

```

RL J. Biol. Chem. 256:12228-12238(1981).
RN [5]
RP SUBCELLULAR LOCATION.
RC TISSUE=LIVER;
RX MEDLINE=82214067; PubMed=6123510;
RA Casey C.A., Anderson P.M.;
RT "Subcellular location of glutamine synthetase and urea cycle enzymes
RT in liver of spiny dogfish (Squalus acanthias).";
RL J. Biol. Chem. 257:8449-8453(1982).
CC -!- FUNCTION: INVOLVED IN THE UREA CYCLE OF UROTELIC ANIMALS WHERE
CC THE ENZYME PLAYS AN IMPORTANT ROLE IN THE GLUTAMINE-DEPENDENT
CC FORMATION OF CARBAMOYL PHOSPHATE FOR UREA SYNTHESIS. UREA MAYBE
CC RETAINED IN TISSUES AND BODY FLUIDS AS A MECHANISM OF
CC OSMOREGULATION.
CC -!- CATALYTIC ACTIVITY: 2 ATP + GLUTAMINE + CO(2) + H(2)O = 2 ADP +
CC PHOSPHATE + GLUTAMATE + CARBAMOYL PHOSPHATE.
CC -!- ENZYME REGULATION: REQUIRES N-ACETYLGLUTAMATE AS AN ALLOSTERIC
CC ACTIVATOR.
CC -!- SUBUNIT: MONOMER.
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN LIVER. MAY NOT BE
CC EXPRESSED IN OTHER TISSUES.
CC -!- SIMILARITY: TO OTHER CARBAMOYL-PHOSPHATE SYNTHETASES. ALSO
CC CONTAINS A GLUTAMINE AMIDOTRANSFERASE-LIKE DOMAIN. THE CYSTEINE
CC RESIDUE WHICH IS ESSENTIAL FOR AMIDO TRANSFERASE ACTIVITY, IS
CC CONSERVED.
DR EMBL; L31362; AAA96435.1; -.
DR EMBL; U19771; AAB49032.1; -.
DR HSSP; P00968; LCS0.
DR InterPro; IPR006275; CarA_L_glu.
DR InterPro; IPR006274; CarA_small.
DR InterPro; IPR005483; CPase_L.
DR InterPro; IPR005479; CPase_L_D2.
DR InterPro; IPR005480; CPase_L_D3.
DR InterPro; IPR005481; CPase_L_N.
DR InterPro; IPR002474; CPase_sm_chain.
DR InterPro; IPR001317; CPS_GATase.
DR InterPro; IPR000991; GATase.1.
DR InterPro; IPR004362; MGS_like.
DR Pfam; PF00289; CPase_L_chain; 2.
DR Pfam; PF02786; CPase_L_D2; 2.
DR Pfam; PF02787; CPase_L_D3; 1.
DR Pfam; PF00988; CPase_sm_chain; 1.
DR Pfam; PF00117; GATase; 1.
DR Pfam; PF02142; MGS; 1.
DR PRINTS; PR00099; CPASASE.
DR PRINTS; PR00099; CPASGATASE.
DR PRINTS; PR00096; GATASE.
DR TIGRFAMS; TIGR01368; CPaseIIsmall; 1.
DR TIGRFAMS; TIGR01369; CPaseII_lrg; 1.
DR PROSITE; PS00866; CPASE_1; 2.
DR PROSITE; PS00867; CPASE_2; 2.
KW Ligase; Translt peptide; Mitochondrion; ATP-binding; Urea cycle.
FT TRANSIT 1 38 MITOCHONDRION.
FT CHAIN 39 1502 CARBAMOYL PHOSPHATE SYNTHETASE III
FT FT [GLUTAMINE] MITOCHONDRIAL.
FT FT ANTHEANILATE PHOSPHORIBOSYLTRANSFERASE-
FT FT LIKE.
FT DOMAIN 39 219 GLUTAMINE AMIDOTRANSFERASE-LIKE.
FT DOMAIN 220 411 CARBAMOYL-PHOSPHATE SYNTHETASE LARGE
FT DOMAIN 412 1502 CHAIN.
FT FT NP_BIND 572 627 ATP (2) (POTENTIAL).
FT NP_BIND 719 769 ATP (1) (POTENTIAL).
FT NP_BIND 1114 1172 ATP (2) (POTENTIAL).
FT NP_BIND 1259 1303 ATP (1) (POTENTIAL).
FT REPEAT 420 877
FT REPEAT 971 1411
FT ACT_SITE 294 294 GATASE (BY SIMILARITY).
SQ SEQUENCE 1502 AA; 165140 MW; CF444C22D99E2EE CRC64;

Query Match      59.9%; Score 163; DB 13; Length 1502;
Best Local Similarity 56.9%; Pred. No. 1.5e-11;
Matches 29; Conservative 7; Mismatches 15; Indels 0; Gaps 0;
```

```

DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Carbamoyl-phosphate synthetase III (EC 6.3.5.5).
OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX NCBI_TaxID=8022;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97197794; PubMed=9045644;
RA Korte J.J., Salo W.L., Cabrera V.M., Wright P.A., Pelskie A.K.,
RA Anderson P.M.;
RT "Expression of carbamoyl-phosphate synthetase III mRNA during the
RT early stages of development and in muscle of adult rainbow trout
RT (Oncorhynchus mykiss).";
RL J. Biol. Chem. 272:6270-6277(1997).
DR EMBL; U65893; AAC60207.1; -.
DR HSSP; P00968; 1CS0.
DR InterPro; IPR006275; CarA_L_glu.
DR InterPro; IPR006274; CarA_small.
DR InterPro; IPR005483; Cphase_L.
DR InterPro; IPR005479; Cphase_L_D2.
DR InterPro; IPR005480; Cphase_L_D3.
DR InterPro; IPR005481; Cphase_L_N.
DR InterPro; IPR002474; CPSase_sm_chain.
DR InterPro; IPR001317; CPS_GATase.
DR InterPro; IPR000991; GATase_L.
DR InterPro; IPR004362; MGS_like.
DR Pfam; PF00289; CPSase_L_chain; 2.
DR Pfam; PF02786; CPSase_L_D2; 2.
DR Pfam; PF02787; CPSase_L_D3; 1.
DR Pfam; PF00988; CPSase_sm_chain; 1.
DR Pfam; PF00117; GATase; 1.
DR Pfam; PF02142; MGS; 1.
DR PRINTS; PR00098; CPSASE.
DR PRINTS; PR00099; CPSGATASE.
DR PRINTS; PR00096; GATASE.
DR TIGRFAMs; TIGR01368; CPSaseIIsmall; 1.
DR TIGRFAMs; TIGR01369; CPSaseII_lrg; 1.
DR PROSITE; PS00866; CPSASE_1; 1.
DR PROSITE; PS00867; CPSASE_2; 2.
KW Ligase.
SQ SEQUENCE 1518 AA; 166577 MW; F9D6331C710A0EBD CRC64;

Query Match 69.9%; Score 190; DB 13; Length 1518;
Best Local Similarity 72.5%; Pred. No. 5.9e-15;
Matches 37; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Oy 1 QLNHGFKLFATEATSDWLNANNVPANVPVPSQGGQNSLSIRKLIRDG 51
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1378 QLNHGFKLFATEATSAWLCANDVPATVPVPSQGGGDSLPSIRLINEG 1428

RESULT 3
O42433
ID 042433 PRELIMINARY; PRT; 1506 AA.
AC 042433;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DE Carbamoyl-phosphate synthetase III (EC 6.3.5.5).
OS Micropterus salmoides (Largemouth bass).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Perciformes; Percoidae;
OC Centrarchidae; Micropterus.
OX NCBI_TaxID=27706;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RA Kong H., Edberg D.D., Salo W.L., Korte J.J., Wright P.A.,
RA Anderson P.M.;

Query Match 62.9%; Score 171; DB 13; Length 1506;
Best Local Similarity 64.0%; Pred. No. 1.5e-12;
Matches 32; Conservative 9; Mismatches 9; Indels 0; Gaps 0;

Oy 2 LHNHGFKLFATEATSDWLNANNVPANVPVPSQGGQNSLSIRKLIRDG 51
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1382 LKRGFKLYATEATSAWLCANDVPATVPVPSQGGGDSLPSIRLINEG 1431

RESULT 4
Q9W7F3
ID Q9W7F3 PRELIMINARY; PRT; 1505 AA.
AC Q9W7F3;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Carbamoyl-phosphate synthetase III.
OS Alcolapia grahami.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Perciformes; Labroidae;
OC Cichlidae; Alcolapia.
OX NCBI_TaxID=87886;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9445540; PubMed=10514466;
RA Lindley T.E., Scheiderer C.L., Walsh P.J., Wood C.M., Bergman H.L.,
RA Bergman A.L., Laurent P., Willson P., Anderson P.M.;
RT "Muscle as the primary site of urea cycle enzyme activity in an
RT alkaline lake-adapted tilapia, Oreochromis alcalicus grahami.";
RL J. Biol. Chem. 274:29858-29861(1999).
DR EMBL; AF119250; AAD43968.1; -.
DR HSSP; P00968; 1A9X.
DR InterPro; IPR006275; CarA_L_glu.
DR InterPro; IPR006274; CarA_small.
DR InterPro; IPR005483; Cphase_L.
DR InterPro; IPR005479; Cphase_L_D2.
DR InterPro; IPR005480; Cphase_L_D3.
DR InterPro; IPR005481; Cphase_L_N.
DR InterPro; IPR002474; CPSase_sm_chain.
DR InterPro; IPR001317; CPS_GATase.

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2003, 02:07:12 ; Search time 57.5 Seconds
(without alignments)
238.881 Million cell updates

Title: US-09-585-077C-2_COPY_1380_1430

Perfect score: 272
Sequence: 1 OLHNEGFKLTATEATSDWLN.....PSOEGONPSLSSIRKLIRDG 51

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_23:*

1: 2: sp_arctaea:*
2: 3: sp_bacteria:*
3: 4: sp_fungi:*
4: 5: sp_human:*
5: 6: sp_invertebrate:*
6: 7: sp_mammal:*
7: 8: sp_mhc:*
8: 9: sp_organelle:*
9: 10: sp_phase:*
10: 11: sp_plant:*
11: 12: sp_rodent:*
12: 13: sp_virus:*
13: 14: sp_vertebrate:*
14: 15: sp_unclassified:*
15: 16: sp_rvms:*
16: 17: sp_bacteriap:*
17: 18: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	260	97.8	748	11	Q8C196	mus musculus
2	190	69.9	1518	13	P70087	oncorhynch
3	171	62.9	1506	13	Q42433	micropterus
4	164	60.3	1505	13	Q9W7F3	alcolapia g
5	163	59.9	1502	13	Q92115	squalus aca
6	156.5	57.5	1504	13	Q9PMN9	psalmus bet
7	80	29.4	2198	5	Q18990	opsanus bet
8	69	25.4	2275	3	Q93937	caenorhabdi
9	67.5	24.8	344	16	Q916H0	psuedomonas
10	64	23.5	863	16	Q8U358	agrobaacteri
11	63.5	23.3	339	13	Q9W709	pleuronecte
12	63.5	23.3	339	13	Q981I8	pleuronecte
13	63.5	23.3	339	13	Q9PW00	pleuronecte
14	63.5	23.3	1798	4	Q14997	rattus norv
15	63.5	23.3	1803	11	Q9ESV1	rattus norv
16	63	23.2	3564	11	Q92313	mus musculu

ALIGNMENTS

RESULT 1

Q8C196
ID Q8C196 PRELIMINARY: PRT: 748 AA.

DT	01-MAR-2003 (TRENBUrel. 23, Created)
DT	01-MAR-2003 (TRENBUrel. 23, Last sequence update)
DT	01-MAR-2003 (TRENBUrel. 23, Last annotation update)
DE	Weakly similar to carbamoyl-phosphate synthase (fragment).
DE	Wus musculus (Mouse).
OS	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC	Mammalia; Butheria; Rodentia; Sciurognathi; Muridae; Mus.
OC	NCRF_TaxID=10090;
ON	[1]
ON	SEQUENCE FROM N.A.
RP	STRAIN=C57BL/6J; TISSUE=Skin;
RC	MEDLINE=22354663; PubMed=12466851;
FX	The FANTOM Consortium,
RA	"Analysis of the mouse transcriptome based on functional annotation of
RT	60,770 full-length cDNAs."
RT	Nature 420:563-573(2002).
RL	EMBL; AK028683; BAC26064.1; -.
RL	NON_TER
FT	SEQUENCE 748 aa, 83277 wgt. AA818F3BA882D9A1 C8C6A.
SO	

QV 1 QLNHNEGFKLFATEATSDWLNANNVPANPVAWPSEQONPSLSSIRKLIRDG 51

628 01:ENEGEK1.FATEATSDWI.NANNVPATPVAWPSPQEGONPSI.SSTRKI.TRDG 678

RESULT 2

P70087
ID P70087 PRELIMINARY: PRT: 1518 AA.

DT 01-FEB-1997 (TREMBlrel. 02, Created)

```
CC phosphate + L-glutamate + carbamoyl phosphate.
CC -!- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -!- PATHWAY: Arginine biosynthesis.
CC -!- PATHWAY: Pyrimidine biosynthesis; first step.
CC -!- SUBUNIT: Composed of two chains; the small (or glutamine) chain
CC promotes the hydrolysis of glutamine to ammonia, which is used by
CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By
CC similarity).
CC -!- SIMILARITY: BELONGS TO THE CARB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL583918; CAC30044.1; ALT_INIT.
CC HSP; P00968; LCS0.
CC Leproma; M0536; .; 1.
CC HAMAP; MF_01210; .; 1.
CC InterPro; IPR006275; CarA_L_glu.
CC InterPro; IPR005483; CPase_L.
CC InterPro; IPR005479; CPase_L_D2.
CC InterPro; IPR005480; CPase_L_D3.
CC InterPro; IPR005481; CPase_L_N.
CC InterPro; IPR004362; MGS_like.
CC Pfam; PF00289; CFSase_L_chain; 2.
CC Pfam; PF02786; CFSase_L_D2; 2.
CC Pfam; PF02787; CFSase_L_D3; 1.
CC Pfam; PF02142; MGS; 1.
CC PRINTS; PR00098; CFSASE.
CC TIGRPFAMs; TIGR01369; CFSaseII_lrg; 1.
CC PROSITE; PS00866; CFSASE_1; 1.
CC PROSITE; PS00867; CFSASE_2; 2.
CC Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
CC ATP-binding; Manganese; Complete proteome.
CC FT DOMAIN 1 407 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
CC FT DOMAIN 408 559 OLIGOMERIZATION DOMAIN.
CC FT DOMAIN 560 965 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
CC FT DOMAIN 966 1121 ALLOSTERIC DOMAIN.
CC FT REPEAT 1 559
CC FT REPEAT 560 1121
CC FT NP_BIND 158 215 ATP (POTENTIAL).
CC FT NP_BIND 308 358 ATP (POTENTIAL).
CC FT METAL 290 290 MANGANESE 1 (BY SIMILARITY).
CC FT METAL 304 304 MANGANESE 1 AND 2 (BY SIMILARITY).
CC FT METAL 306 306 MANGANESE 2 (BY SIMILARITY).
CC FT METAL 843 843 MANGANESE 3 (BY SIMILARITY).
CC FT METAL 855 855 MANGANESE 3 (BY SIMILARITY).
CC SQ SEQUENCE 1121 AA; 120016 MW; 2EABCBA0A8D4281CD CRC64;
Query Match 21.5%; Score 58.5; DB 1; Length 1121;
Best Local Similarity 33.3%; Pred. No. 29;
Matches 18; Conservative 12; Mismatches 19; Indels 5; Gaps 3;
Oy 1 QLHNEGKLFATEATSDWLNANNVPANPVA---WPQEGONPSLSIRKLIRDG 51
Db 992 RLADLGFVIVTGTCTAEMLRKGFCDVVRKHFEPPKAGR-PALSAV-DAIRAG 1043
```

Search completed: September 13, 2003, 02:13:28
Job time : 18 secs

XX	MEDLINE=95282517; PubMed=7762305;	RT	"Massive gene decay in the leprosy bacillus.";
AT	Coster F., van Dyck L., Joniaux J., Purnelle B., Coffeau A.;	RL	Nature 409:1007-1011(2001).
TA	"The sequence of a 13.5 kb DNA segment from the left arm of yeast	CC	CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + Co(2) + H(2)O = 2 ADP +

DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE URAL protein [includes: Glutamine-dependent carbamoyl-phosphate
synthase (EC 6.3.5.5); Aspartate carbamoyltransferase (EC 2.1.3.2)].
GN URAL OR SPAC2267.06C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE OF 22-2244 FROM N.A.
RC STRAIN=972;
RX MEDLINE=96020160; PubMed=8590465;
RA Lollier M., Jaquet L., Nedeva T., Lacroute F., Potier S.,
RA Souciet J.-L.;
RT "As in Saccharomyces cerevisiae, aspartate transcarbamoylase is
RT assembled on a multifunctional protein including a dihydroorotase-like
RT cryptic domain in Schizosaccharomyces pombe.";
RL Curr. Genet. 28:138-149(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgourou J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Church C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feldwell T., Fraser A.,
RA Gattiles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Money P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart E., Aert R., Robben J., Grynponez B.,
RA Weltjens I., Vantreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsberg S.L.,
RA Shpakovskii G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -!- FUNCTION: THIS PROTEIN IS A "FUSION" PROTEIN ENCODING THREE
CC ENZYMAIC ACTIVITIES OF THE PYRIMIDINE PATHWAY (GATASE, CPSASE,
CC AND ATCASE) (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
CC phosphate + L-glutamate + carbamoyl phosphate.
CC -!- CATALYTIC ACTIVITY: Carbamoyl phosphate + L-aspartate = phosphate
CC + N-carbamoyl-L-aspartate.
CC -!- PATHWAY: Pyrimidine biosynthesis; first step.
CC -!- PATHWAY: Pyrimidine biosynthesis; second step.
CC -!- MISCELLANEOUS: GATASE (GLUTAMINE AMIDOTRANSFERASE) AND CPSASE
CC (CARBAMOYL PHOSPHATE SYNTHASE) FORM TOGETHER THE
CC GLUTAMINE-DEPENDENT CPSASE (GD-CPSASE) (EC 6.3.5.5).
CC -!- MISCELLANEOUS: IN EUKARYOTES EC 6.3.5.5 IS SYNTHESIZED BY TWO
CC PATHWAY-SPECIFIC (ARGININE AND PYRIMIDINE) UNDER SEPARATE CONTROL.
CC -!- SIMILARITY: THE CPSASE DOMAIN IS SIMILAR TO OTHER CPSASES.
CC -!- SIMILARITY: IN THE CENTRAL SECTION; BELONGS TO THE DHOASE FAMILY.
CC DEFECTIVE DOMAIN.
CC -!- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X81841; CAA57433.1; -;
CC EMBL; Z54328; CAA91130.1; -;
CC PIR; S65074; S65074.
CC PIR; T11616; T11616.
CC HSSP; P00968; 1A9X.
CC GeneDB_SPombe; SPAC22G7.06C; -;
CC InterPro; IPR006220; Anth_synthII.
CC InterPro; IPR006130; Asp/Orn_Cotransf.
CC InterPro; IPR002082; Asp_carbmltransf.
CC InterPro; IPR006275; CarA_L_glu.
CC InterPro; IPR006274; CarA_small.
CC InterPro; IPR005483; CPase_I.
CC InterPro; IPR005479; CPase_L_D2.
CC InterPro; IPR005480; CPase_L_D3.
CC InterPro; IPR005481; CPase_L_N.
CC InterPro; IPR001317; CPS_GATase.
CC InterPro; IPR002474; CPSase_sm_chain.
CC InterPro; IPR000991; GATase_1.
CC InterPro; IPR004362; MGS_like.
CC InterPro; IPR006131; OTCace_O.
CC InterPro; IPR006132; OTCace_P.
CC Pfam; PF00289; CPsase_L_chain; 2.
CC Pfam; PF02786; CPsase_L_D2; 2.
CC Pfam; PF02787; CPsase_L_D3; 1.
CC Pfam; PF00988; CPsase_sm_chain; 1.
CC Pfam; PF02117; GATase; 1.
CC Pfam; PF02142; MGS; 1.
CC Pfam; PF00185; OTCace; 1.
CC Pfam; PF02729; OTCace_N; 1.
CC PRINTS; PR00097; ANTSENTHASEII.
CC PRINTS; PR00100; ACTCASE.
CC PRINTS; PR00098; CPSASE.
CC PRINTS; PR00099; CPSGATASE.
CC PRINTS; PR00096; GATASE.
CC TIGRFAMS; TIGR00670; asp_carb_tr; 1.
CC TIGRFAMS; TIGR01369; CPsaseII_lrg; 1.
CC TIGRFAMS; TIGR01368; CPsaseII_small; 1.
CC PROSITE; PS00097; CARBAMOYLTRANSFERASE; 1.
CC PROSITE; PS00442; GATASE_TYPE_I; 1.
CC PROSITE; PS00866; CPSASE_1; 2.
CC PROSITE; PS00867; CPSASE_2; 2.
CC Pyrimidine biosynthesis; Ligase; Transferase; Multifunctional enzyme.
CC GATASE (GLUTAMINE AMIDOTRANSFERASE)
CC (BY SIMILARITY).
CC LINKER (BY SIMILARITY).
CC CPSASE (CARBAMOYL-PHOSPHATE SYNTHASE)
CC (BY SIMILARITY).
CC LINKER (BY SIMILARITY).
CC DEFECTIVE DHOASE DOMAIN
CC (BY SIMILARITY).
CC LINKER (BY SIMILARITY).
CC ATCASE (ASPARTATE TRANSCARBAMYLASE)
CC (BY SIMILARITY).
CC GATASE (BY SIMILARITY).
CC GATASE (BY SIMILARITY).
CC GATASE (BY SIMILARITY).
CC GIC -> RVF (IN REF. 1).
CC CAVRA -> LQFAQ (IN REF. 1).
CC EL -> DV (IN REF. 1).
CC G -> E (IN REF. 1).
CC G -> E (IN REF. 1).
CC CONFLICT 2002 2002
CC CONFLICT 2002 2002
CC SEQUENCE 2244 AA; 248306 MW; 5700D153B50CD3E9 CRC64;

CC Query Match 23.5%; Score 64; DB 1; Length 2244;
CC Best Local Similarity 35.6%; Pred. No. 13;
CC Matches 16; Conservative 10; Mismatches 17; Indels 2; Gaps 2;

CC QY 1 QLHNEGFKLFATEATSDMLNANNYPANPVA-WFSQEGQNP-SLSS 43

Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S., Seeger K., Saunders D., Sharp S., Squares S., Taylor K., Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J., Hopwood D.A.;
"Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2)".
Nature 417:141-147(2002).
-|- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP + phosphate + L-glutamate + carbamoyl phosphate.
-|- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
-|- PATHWAY: Arginine biosynthesis.
-|- PATHWAY: Pyrimidine biosynthesis; first step.
-|- SUBUNIT: Composed of two chains; the small (or glutamine) chain promotes the hydrolysis of glutamine to ammonia, which is used by the large (or ammonia) chain to synthesize carbamoyl phosphate (By similarity).
-|- SIMILARITY: BELONGS TO THE CARB FAMILY.
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation on its European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
EMBL; AL939109; CAB93363.1; -.
HSSP; P00968; 1CSO.
HAMAP; MF_01210; -. 1.
InterPro; IPR006275; CarA_L_glu.
InterPro; IPR005483; Cphase_1.
InterPro; IPR005479; Cphase_1_D2.
InterPro; IPR005480; Cphase_1_D3.
InterPro; IPR005481; Cphase_1_N.
InterPro; IPR004352; MGS_like.
Pfam; PF00289; CPhase_L_chain; 2.
Pfam; PF02786; CPhase_L_D2; 3.
Pfam; PF02787; CPhase_L_D3; 1.
Pfam; PF02142; MGS; 1.
PRINTS; PR00098; CPSASE.
TIGRFS; TIGR01369; CPSaseII_lig; 1.
PROSITE; PS00856; CPSASE_1; 2.
PROSITE; PS00867; CPSASE_2; 2.
Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat; ATP-binding; Manganese; Complete proteome.
KW CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 1 408 OLIGOMERIZATION DOMAIN.
FT DOMAIN 409 553 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 554 954 ALLOSTERIC DOMAIN.
FT DOMAIN 955 1102 ALLOSTERIC DOMAIN.
FT REPEAT 1 553
FT REPEAT 554 1102
FT REPEAT 158 216 ATP (POTENTIAL).
FT NP_BIND 309 359 ATP (POTENTIAL).
FT FT METAL 291 291 MANGANESE 1 (BY SIMILARITY).
FT FT METAL 305 305 MANGANESE 1 AND 2 (BY SIMILARITY).
FT FT METAL 307 307 MANGANESE 2 (BY SIMILARITY).
FT FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).
FT FT METAL 844 844 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1102 AA; 117908 MW; E72DA263C988DE02 CRC64;
Query Match 23.5%; Score 64; DB 1; Length 1102;
Best Local Similarity 38.5%; Pred No. 5.9;
Matches 20; Conservative 8; Mismatches 22; Indels 2; Gaps 2;
QY 1 QLNHEGFKLFATPTSDWLNANNVPANVPANVPSQEQNPS-LSSIRKLIRDG 51
: ||| ||| :
Db 961 ELYAHGFELMATSGAEVLKRGINATVVRKQS-EQTGPNCEKTIIVQLHDG 1031
RESULT 12
PYEL_SCHPO
ID PYEL_SCHPO STANDARD; PRT; 2244 AA.
AC Q09794;

[illegible]


```

DR Pfam; PF02142; MGS; 1.
DR PRINTS; PR00098; CPSASE.
DR TIGRFAMS; TIGR01369; CPSaseII_lrg; 1.
DR PROSITE; PS00866; CPSASE.1; 2.
DR PROSITE; PS00867; CPSASE.2; 2.
KW Pyrimidine biosynthesis; Ligase; Repeat; ATP-binding; Manganese;
KW Complete proteome.
FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 402 546 OLIGOMERIZATION DOMAIN.
FT DOMAIN 547 929 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 930 1071 ALLOSTERIC DOMAIN.
FT REPEAT 1 546
FT REPEAT 547 1071
FT NP_BIND 153 210 ATP (POTENTIAL).
FT NP_BIND 302 332 ATP (POTENTIAL).
FT METAL 284 284 MANGANESE 1 (BY SIMILARITY).
FT METAL 298 298 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 300 300 MANGANESE 2 (BY SIMILARITY).
FT METAL 820 820 MANGANESE 3 (BY SIMILARITY).
FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1071 AA; 117649 MW; 195E33CFC5C6222C CRC64;

Query Match 25.6%; Score 69.5; DB 1; Length 1071;
Best Local Similarity 31.4%; Pred. No. 1.2;
Matches 16; Conservative 11; Mismatches 19; Indels 5; Gaps 1;

QY 1 QLNHEGKFLFAFETSDMNNVNPANVPVQSGNQPSLSIRKLIRDG 51
: | : ||| : | : ||| : | : ||| : | : ||| : | : ||| :
DB 956 RFAIGNYLATEAGTGYLKEASIPAKVVGKIGDGFN-----LLDVIRNG 1001

RESULT 7
PYRLDICI STANDARD; PRT; 2185 AA.
AC P20054;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protein PYRL-3 [Includes: Glutamine-dependent carbamoyl-phosphate
DE synthase (EC 6.3.5.5); Aspartate carbamoyltransferase (EC 2.1.3.2);
DE Dihydroorotase (EC 3.5.2.3)].
GN PYRL-3.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE OF 1-457 AND 1175-2185 FROM N.A.
RC STRAIN=AX3;
RC MEDLINE=89137111; PubMed=2917570;
RA Faure M., Camonis J.H., Jacquet M.;
RT "Molecular characterization of a Dictyostelium discoideum gene
RT encoding a multifunctional enzyme of the pyrimidine pathway.";
RL Eur. J. Biochem. 179:345-356(1989).
RN [2]
RP SEQUENCE OF 367-1408 FROM N.A.
RC STRAIN=AX2;
RC MEDLINE=92329976; PubMed=1627825;
RA Elgar G., Schofield J.P.;
RT "Carbamoyl phosphate synthetase (CPSase) in the PYRL-3 multigene of
RT Dictyostelium discoideum.";
RL DNA Seq. 2:219-226(1992).
CC -!- FUNCTION: THIS PROTEIN IS A "FUSION" PROTEIN ENCODING FOUR
CC ENZYMIC ACTIVITIES OF THE PYRIMIDINE PATHWAY (GATASE, CPSASE,
CC ATCASE AND DHASE).
CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
CC phosphate + L-glutamate + carbamoyl phosphate.
CC -!- CATALYTIC ACTIVITY: Carbamoyl phosphate + L-aspartate = phosphate
CC + N-carbamoyl-L-aspartate.
CC -!- CATALYTIC ACTIVITY: (S)-dihydroorotate + H(2)O = N-carbamoyl-L-
CC aspartate.
CC -!- COFACTOR: DHASE ACTIVITY REQUIRES A ZINC ATOM.
CC -!- ENZYME REGULATION: ALLOSTERICALLY REGULATED AND CONTROLLED BY
CC PHOSPHORYLATION. PRPP IS AN ACTIVATOR WHILE UMP IS AN INHIBITOR

```

```

CC OF THE CPSASE REACTION.
CC -!- PATHWAY: FIRST TO THIRD STEP IN PYRIMIDINE BIOSYNTHESIS.
CC -!- SUBUNIT: Homohexamer.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- DEVELOPMENTAL STAGE: SEEN DURING GROWTH BUT NOT DURING
CC DEVELOPMENT.
CC -!- MISCELLANEOUS: GATASE (GLUTAMINE AMIDOTRANSFERASE) AND CPSASE
CC (CARBAMOYL PHOSPHATE SYNTHASE) FORM TOGETHER THE
CC GLUTAMINE-DEPENDENT CPSASE (GD-CPSASE) (EC 6.3.5.5).
CC -!- SIMILARITY: THE CPSASE DOMAIN IS SIMILAR TO OTHER CPSASES.
CC -!- SIMILARITY: IN THE CENTRAL SECTION; BELONGS TO THE DHASE FAMILY.
CC -!- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X14633; CRA32781.1; .
CC EMBL; X14634; CRA32782.1; .
CC EMBL; X55433; CRA39077.1; .
CC PIR; S23738; S23738.
CC HSP; P00479; 3CSU.
CC DictyDB; DD05019; pyr1-3.
CC InterPro; IPR006680; Amidohydro_1.
CC InterPro; IPR006130; Asp/Orn_Cotransf.
CC InterPro; IPR002082; Asp_carbamyltransf.
CC InterPro; IPR006275; CarA_L_glu.
CC InterPro; IPR006274; CarA_small.
CC InterPro; IPR005483; CPase_L.
CC InterPro; IPR005479; CPase_L_D2.
CC InterPro; IPR005480; CPase_L_D3.
CC InterPro; IPR005481; CPase_L_N.
CC InterPro; IPR001317; CPS_Gatase.
CC InterPro; IPR002474; CPase_sm_chain.
CC InterPro; IPR002195; Dihydroorotase.
CC InterPro; IPR000991; Gatase_L.
CC InterPro; IPR004362; MGS_like.
CC InterPro; IPR006131; OTCace_O.
CC InterPro; IPR006132; OTCace_P.
CC InterPro; IPR004722; PyrC_multi.
CC InterPro; IPR005847; Urease.
CC Pfam; PF01979; Amidohydro_L; 1.
CC Pfam; PF00289; CPSase_L_chain; 2.
CC Pfam; PF02786; CPSase_L_D2; 2.
CC Pfam; PF02787; CPSase_L_D3; 1.
CC Pfam; PF00988; CPSase_sm_chain; 1.
CC Pfam; PF00117; GATase; 1.
CC Pfam; PF02142; MGS; 1.
CC Pfam; PF00185; OTCace; 1.
CC Pfam; PF02729; OTCace_N; 1.
CC PRINTS; PR00100; AOTCASE.
CC PRINTS; PR00098; CPSASE.
CC PRINTS; PR00099; CPSGATASE.
CC PRINTS; PR00096; GATASE.
CC ProDom; PD000518; Urease; 1.
CC TIGRFAMS; TIGR00670; asp_carb_tr; 1.
CC TIGRFAMS; TIGR01369; CPSaseII_lrg; 1.
CC TIGRFAMS; TIGR01368; CPSaseI_small; 1.
CC TIGRFAMS; TIGR00857; pyrC_multi; 1.
CC PROSITE; PS00097; CARBAMOYLTRANSFERASE; 1.
CC PROSITE; PS00442; GATASE_TYPE_I; 1.
CC PROSITE; PS00482; DIHYDROOROTASE_1; 1.
CC PROSITE; PS00483; DIHYDROOROTASE_2; 1.
CC PROSITE; PS00866; CPSASE_1; 2.
CC PROSITE; PS00867; CPSASE_2; 2.
KW Pyrimidine biosynthesis; Ligase; Transferase; Hydrolase;
KW Allosteric enzyme; Zinc; Multifunctional enzyme.
FT DOMAIN 1 351 GATASE (GLUTAMINE AMIDOTRANSFERASE).
FT DOMAIN 352 366 LINKER.

```

```

CC CC similarity).
CC CC -!- SIMILARITY: BELONGS TO THE CARB FAMILY.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; Z81011; ; NOT_ANNOTATED.CDS.
CC CC EMBL; Z80108; ; NOT_ANNOTATED.CDS.
CC CC EMBL; AE007014; AK43593.1; -.
CC CC PIR; A70990; A70990.
CC CC HSSP; P00968; ICS0.
CC CC TIGR; MT1428; -.
CC CC TuberculList; Rv1384; -.
CC CC HAMAP; MF_01210; ; 1.
CC CC InterPro; IPR006275; CarA_L_glu.
CC CC InterPro; IPR005483; CPase_L.
CC CC InterPro; IPR005479; CPase_L_D2.
CC CC InterPro; IPR005480; CPase_L_D3.
CC CC InterPro; IPR005481; CPase_L_N.
CC CC InterPro; IPR004362; MGS_like.
CC CC Pfam; PF00289; CPase_L_chain; 2.
CC CC Pfam; PF02786; CPase_L_D2; 3.
CC CC Pfam; PF02142; MGS; 1.
CC CC PRINTS; PR00098; CPASE.
CC CC TIGRFAMs; TIGR01369; CPaseII_lrg; 1.
CC CC PROSITE; PS00866; CPASE_1; 1.
CC CC PROSITE; PS00867; CPASE_2; 2.
CC CC Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
CC CC ATP-binding; Manganese; Complete proteome.
CC CC FT DOMAIN 1 407 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
CC CC FT DOMAIN 408 559 OLIGOMERIZATION DOMAIN.
CC CC FT DOMAIN 560 965 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
CC CC FT DOMAIN 966 1115 ALLOSTERIC DOMAIN.
CC CC FT REPEAT 1 559
CC CC FT REPEAT 560 1115
CC CC FT NP_BIND 158 215 ATP (POTENTIAL).
CC CC FT NP_BIND 308 358 ATP (POTENTIAL).
CC CC FT METAL 290 290 MANGANESE 1 (BY SIMILARITY).
CC CC FT METAL 304 304 MANGANESE 1 AND 2 (BY SIMILARITY).
CC CC FT METAL 306 306 MANGANESE 2 (BY SIMILARITY).
CC CC FT METAL 843 843 MANGANESE 3 (BY SIMILARITY).
CC CC FT METAL 855 855 MANGANESE 3 (BY SIMILARITY).
CC CC SQ SEQUENCE 1115 AA; 118961 MW; 293F6A35FDE05981 CRC64;
CC CC Query Match 26.3%; Score 71.5; DB 1; Length 1115;
CC CC Best Local Similarity 35.2%; Pred. No. 0.69;
CC CC Matches 19; Conservative 14; Mismatches 16; Indels 5; Gaps 3;
CC CC QY 1 QHNEGKFLKATEATSDWLNANNVPANVA---WPSEGONPSLSIRKLIRDG 51
CC CC Db 992 RLADLGEFLVLTACTAEMLRNGICDDYRKHFPEPAQGR-PTMSAV-DATAG 1043
CC CC
CC CC RESULT 6
CC CC CARB_BACSU
CC CC ID CARB_BACSU STANDARD; PRT; 1071 AA.
CC CC AC P25994;
CC CC DT 01-MAY-1992 (Rel. 22, Created)
CC CC DT 01-MAY-1992 (Rel. 22, Last sequence update)
CC CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC CC DE Carbamoyl-phosphate synthase, pyrimidine-specific, large chain
CC CC DE (EC 6.3.5.5) (Carbamoyl-phosphate synthetase ammonia chain).
CC CC GN PRAB.
CC CC OS Bacillus subtilis.
CC CC OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
CC CC OX NCBI_TaxID=1423;
CC CC EN [1]

```

```

RP RX SEQUENCE FROM N.A.
RX MEDLINE=91225016; PubMed=1709162;
RA Quinn C.L., Stephenson B.T., Swinn R.L.;
RT "Functional organization and nucleotide sequence of the Bacillus
RL J. Biol. Chem. 266:9113-9127(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Chim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.P., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinio S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Lin H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogilwa A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scalan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
FT subtilis.";
RL Nature 390:249-256(1997).
CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
CC phosphate + L-glutamate + carbamoyl phosphate.
CC -!- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -!- PATHWAY: Pyrimidine biosynthesis; first step.
CC -!- SUBUNIT: Composed of two chains; the small (or glutamine) chain
CC promotes the hydrolysis of glutamine to ammonia, which is used by
CC the large (or ammonia) chain to synthesize carbamoyl phosphate.
CC -!- SIMILARITY: BELONGS TO THE CARB FAMILY.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; M59757; AAA21270.1; -.
CC CC EMBL; Z99112; CAB13426.1; -.
CC CC PIR; F39845; F39845.
CC CC HSSP; P00968; 1JDB.
CC CC Subtilist; BGI0716; pyRAB.
CC CC HAMAP; MF_01210; ; 1.
CC CC InterPro; IPR006275; CarA_L_glu.
CC CC InterPro; IPR005483; CPase_L.
CC CC InterPro; IPR005479; CPase_L_D2.
CC CC InterPro; IPR005480; CPase_L_D3.
CC CC InterPro; IPR005481; CPase_L_N.
CC CC InterPro; IPR004362; MGS_like.
CC CC Pfam; PF00289; CPase_L_chain; 2.
CC CC Pfam; PF02786; CPase_L_D2; 2.
CC CC Pfam; PF02787; CPase_L_D3; 1.

```

```
FT REPEAT 416 873
FT REPEAT 967 1407
FT ACT SITE 290 290 GATASE (BY SIMILARITY).
SQ SEQUENCE 1496 AA; 163767 MW; 2360BF05D21B059E CRC64;

Query Match 66.0%; Score 179.5; DB 1; Length 1496;
Best Local Similarity 70.0%; Pred. No. 3.7e-14;
Matches 33; Conservative 8; Mismatches 6; Indels 1; Gaps 1;

QY 2 LHNEGFKLFATEATSDWLNANNVPANVPWPSQEQNPSLSIRKLIRDG 51
   I :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1378 LKDEGFKLFATEATSDWLNANDITATVPWPSQEQG-CPSSIVKLIKEG 1426

RESULT 4
CAB2_METJA
ID CAB2_METJA STANDARD; PRT; 618 AA.
AC Q58776;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain, C-terminal section
   (EC 6.3.5.5) (Carbamoyl-phosphate synthetase ammonia chain).
GN CAB2 OR MJ1381.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
   jannaschii."
RL Science 273:1058-1073(1996).
CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
   phosphate + L-glutamate + carbamoyl phosphate.
CC -!- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -!- PATHWAY: Arginine biosynthesis; first step.
CC -!- PATHWAY: Pyrimidine biosynthesis; the small (or glutamine) chain
   promotes the hydrolysis of glutamine to ammonia, which is used by
   the large (or ammonia) chain to synthesize carbamoyl phosphate (By
   similarity).
CC -!- SIMILARITY: BELONGS TO THE CARB FAMILY. C-TERMINAL SECTION.
CC -!- CAUTION: Sequence of carb is split into two genes in M.jannaschii
   (MJ1378 and MJ1381).
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; U67578; AB99391.1; -.
FIR; D64472; D64472.
ESP; P00968; IEXR.
TIGR; MJ1381; -.
DR HAMAP; MF_01210; atypical; 1.
DR InterPro; IPR005483; CPase_L.
DR InterPro; IPR005479; CPase_L_D2.
DR InterPro; IPR005480; CPase_L_D3.
DR InterPro; IPR005481; CPase_L_N.

DR InterPro; IPR004362; MGS_like.
DR Pfam; PF00289; CPase_L_chain; 1.
DR Pfam; PF02786; CPase_L_D2; 2.
DR Pfam; PF02787; CPase_L_D3; 1.
DR Pfam; PF02142; MGS; 1.
DR PRINTS; PR00098; CPASE.
DR PROSITE; PS00866; CPASE_1; 1.
DR PROSITE; PS00867; CPASE_2; 1.
KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; ATP-binding;
   Manganese; Complete proteome.
FT DOMAIN 81 477 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 478 618 ALLOSTERIC DOMAIN.
FT METAL 358 358 MANGANESE 3 (BY SIMILARITY).
FT METAL 370 370 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 618 AA; 68453 MW; 8C3D4D9C192A954E CRC64;

Query Match 29.4%; Score 80; DB 1; Length 618;
Best Local Similarity 39.2%; Pred. No. 0.031;
Matches 20; Conservative 11; Mismatches 16; Indels 4; Gaps 2;

QY 1 QLHNEGFKLFATEATSDWLNANNVPANVPWPSQEQNPSLSIRKLIRDG 51
   :||| ||| :||||| :| | :||| | :||| :||| :||| :|||
Db 502 KLHLEGTFTYATEGAKVIRENGIPALV---KKISESPN-DNLIKLMRDG 548

RESULT 5
CARB_MYCTU
ID CARB_MYCTU STANDARD; PRT; 1115 AA.
AC P57689;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
   phosphate synthetase ammonia chain).
GN CARB OR RV1384 OR MT1428 OR MTCY02B12.18 OR MTCY21B4.01.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E., III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
   complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleisichmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
   laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
   phosphate + L-glutamate + carbamoyl phosphate.
CC -!- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -!- PATHWAY: Arginine biosynthesis; first step.
CC -!- PATHWAY: Pyrimidine biosynthesis; the small (or glutamine) chain
   promotes the hydrolysis of glutamine to ammonia, which is used by
   the large (or ammonia) chain to synthesize carbamoyl phosphate (By
```

```

CC DE EMBL; M12335; AAB59717.1; -.
DR DE EMBL; M11710; AAB59717.1; JOINED.
DR OS EMBL; M12318; AAB59717.1; JOINED.
DR OS EMBL; M12319; AAB59717.1; JOINED.
DR OS EMBL; M12320; AAB59717.1; JOINED.
DR OS EMBL; M12321; AAB59717.1; JOINED.
DR OS EMBL; M12322; AAB59717.1; JOINED.
DR OS EMBL; M12323; AAB59717.1; JOINED.
DR OS EMBL; M12324; AAB59717.1; JOINED.
DR OS EMBL; M12325; AAB59717.1; JOINED.
DR OS EMBL; M12326; AAB59717.1; JOINED.
DR OS EMBL; M12327; AAB59717.1; JOINED.
DR OS EMBL; M12328; AAB59717.1; JOINED.
DR OS EMBL; J02805; AAA40959.1; -.
DR DE PIR; A28481; SYRTCA.
DR DE HSSP; P00968; LCSO.
DR DE InterPro; IPR006275; CarA_L_glu.
DR DE InterPro; IPR006274; CarA_small.
DR DE InterPro; IPR005483; CPase_L.
DR DE InterPro; IPR005479; CPase_L_D2.
DR DE InterPro; IPR005480; CPase_L_D3.
DR DE InterPro; IPR005481; CPase_L_N.
DR DE InterPro; IPR001317; CPS_GATase.
DR DE InterPro; IPR002474; CPase_sm_chain.
DR DE InterPro; IPR000991; GATase_L.
DR DE InterPro; IPR004362; MGS_like.
DR DE Pfam; PF02786; CPase_L_chain; 2.
DR DE Pfam; PF02787; CPase_L_D3; 1.
DR DE Pfam; PF00988; CPase_sm_chain; 1.
DR DE Pfam; PF00117; GATase; 1.
DR DE Pfam; PF02142; MGS; 1.
DR DE PRINTS; PR00098; CPSASE.
DR DE PRINTS; PR00099; GATASE.
DR DE TIGRFAMS; TIGR01369; CPaseII_lrg; 1.
DR DE TIGRFAMS; TIGR01368; CPaseIISmall; 1.
DR DE PROSITE; PS00866; CPSASE_1; 2.
DR DE PROSITE; PS00867; CPSASE_2; 2.
KW Ligase; Allosteric enzyme; Repeat; Transit peptide; Mitochondrion;
FT ATP-binding; Urea cycle.
FT TRANSIT 1 38 MITOCHONDRION.
FT CHAIN 39 1500 CARBAMOYL-PHOSPHATE SYNTHASE [AMMONIA].
FT NP_BIND 571 626 ATP (2).
FT NP_BIND 718 768 ATP (1).
FT NP_BIND 1113 1171 ATP (2).
FT NP_BIND 1259 1302 ATP (1).
FT DOMAIN 39 219 ANTHRANILATE PHOSPHORIBOSYLTRANSFERASE
FT HOMOLOG.
FT DOMAIN 220 410 GLUTAMINE AMIDOTRANSFERASE-LIKE.
FT DOMAIN 411 1500 CARBAMOYL-PHOSPHATE SYNTHETASE LARGE
FT CHAIN.
FT REPEAT 419 876
FT REPEAT 970 1410
FT SEQUENCE 1500 AA; 164579 MW; 038E8F893DE1C34D CRC64;
Query Match 97.8%; Score 266; DB 1; Length 1500;
Best Local Similarity 98.0%; Pred. No. 6.5e-25;
Matches 50; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 QLNHEGKLFATFATSDMLNANNVPANVPVAPVPSQEGNPSLSIRKLRDG 51
DB 1380 QLNHEGKLFATFATSDMLNANNVPANVPVAPVPSQEGNPSLSIRKLRDG 1430
RESULT 3
ID CPM_RANCA STANDARD; PRT; 1496 AA.
AC Q91293;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)

```

Carbamoyl-phosphate synthase [ammonia], mitochondrial precursor (EC 6.3.4.16) (Carbamoyl-phosphate synthetase I) (CPSASE I).
Rana catesbeiana (Bull frog).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Neobatrachia; Ranidae; Rana.
NCBI_TaxID=8400;
[1]
SEQUENCE FROM N.A.
TISUP=Liver;
MEDLINE=94216272; PubMed=8163471;
Helbing C.C., Atkinson B.G.;
"3,5,3'-triiodothyronine-induced carbamyl-phosphate synthetase gene expression is stabilized in the liver of Rana catesbeiana tadpoles during heat shock.";
J. Biol. Chem. 269:11743-11750(1994).
-!- FUNCTION: INVOLVED IN THE UREA CYCLE OF UREOTELIC ANIMALS WHERE THE ENZYME PLAYS AN IMPORTANT ROLE IN REMOVING EXCESS AMMONIA FROM THE CELL (BY SIMILARITY).
-!- CATALYTIC ACTIVITY: 2 ATP + NH(3) + CO(2) + H(2)O = 2 ADP + phosphate + carbamoyl phosphate.
-!- ENZYME REGULATION: REQUIRES N-ACETYLGLUTAMATE AS AN ALLOSTERIC ACTIVATOR (BY SIMILARITY).
-!- SUBCELLULAR LOCATION: Mitochondrial.
-!- SIMILARITY: TO OTHER CARBAMOYL-PHOSPHATE SYNTHETASES. ALSO CONTAINS A GLUTAMINE AMIDOTRANSFERASE-LIKE DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

EMBL; U05193; AAA19016.1; -.
PIR; I51170; I51170.
HSSP; P00968; LCSO.
InterPro; IPR006275; CarA_L_glu.
InterPro; IPR006274; CarA_small.
InterPro; IPR005483; CPase_L.
InterPro; IPR005479; CPase_L_D2.
InterPro; IPR005480; CPase_L_D3.
InterPro; IPR005481; CPase_L_N.
InterPro; IPR001317; CPS_GATase.
InterPro; IPR002474; CPase_sm_chain.
InterPro; IPR000991; GATase_L.
InterPro; IPR004362; MGS_like.
Pfam; PF00289; CPase_L_chain; 2.
Pfam; PF02786; CPase_L_D2; 2.
Pfam; PF02787; CPase_L_D3; 1.
Pfam; PF00988; CPase_sm_chain; 1.
Pfam; PF00117; GATase; 1.
Pfam; PF02142; MGS; 1.
PRINTS; PR00098; CPSASE.
PRINTS; PR00099; GATASE.
TIGRFAMS; TIGR01369; CPaseII_lrg; 1.
TIGRFAMS; TIGR01368; CPaseIISmall; 1.
PROSITE; PS00866; CPSASE_1; 2.
PROSITE; PS00867; CPSASE_2; 2.
Ligase; Allosteric enzyme; Repeat; Transit peptide; Mitochondrion; ATP-binding; Urea cycle.
TRANSIT 1 38 MITOCHONDRION.
CHAIN 39 1500 CARBAMOYL-PHOSPHATE SYNTHASE [AMMONIA].
NP_BIND 571 626 ATP (2).
NP_BIND 718 768 ATP (1).
NP_BIND 1113 1171 ATP (2).
NP_BIND 1259 1302 ATP (1).
DOMAIN 39 219 ANTHRANILATE PHOSPHORIBOSYLTRANSFERASE
HOMOLOG.
DOMAIN 220 410 GLUTAMINE AMIDOTRANSFERASE-LIKE.
DOMAIN 411 1500 CARBAMOYL-PHOSPHATE SYNTHETASE LARGE
CHAIN.
REPEAT 419 876
REPEAT 970 1410
SEQUENCE 1500 AA; 164579 MW; 038E8F893DE1C34D CRC64;
Query Match 97.8%; Score 266; DB 1; Length 1500;
Best Local Similarity 98.0%; Pred. No. 6.5e-25;
Matches 50; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 QLNHEGKLFATFATSDMLNANNVPANVPVAPVPSQEGNPSLSIRKLRDG 51
DB 1380 QLNHEGKLFATFATSDMLNANNVPANVPVAPVPSQEGNPSLSIRKLRDG 1430
RESULT 3
ID CPM_RANCA STANDARD; PRT; 1496 AA.
AC Q91293;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)

CC symptoms are vomiting in infancy, protein intolerance,
 CC intermittent ataxia, seizures, lethargy, and mental retardation.
 CC -!- SIMILARITY: TO OTHER CARBAMOYL-PHOSPHATE SYNTHETASES. ALSO
 CC CONTAINS A GLUTAMINE AMIDOTRANSFERASE-LIKE DOMAIN BUT THE CYSTEINE
 CC RESIDUE WHICH IS ESSENTIAL FOR AMIDO TRANSFERASE ACTIVITY, IS
 CC MISSING IN THIS PROTEIN SO THIS DOMAIN IS NO LONGER ACTIVE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D90282; BAA14328.1; -;
 CC EMBL; Y15793; CAA5785.1; -;
 CC EMBL; AF154830; AAD38072.1; -;
 CC PIR; JQ1348; JQ1348.
 CC HSP; P00968; IA9X.
 CC Genew; HGNC:2323; CPS1.
 CC GK; P31327; -;
 CC MIM; 237300; -;
 CC GO; GO:0005739; C-mitochondrion; TAS.
 CC GO; GO:0004087; F:carbamoyl-phosphate synthase (ammonia) acti. . .; TAS.
 CC InterPro; IPR006275; CarA_L_glu.
 CC InterPro; IPR006274; CarA_small.
 CC InterPro; IPR005483; CPase_L.
 CC InterPro; IPR005479; CPase_L_D2.
 CC InterPro; IPR005480; CPase_L_D3.
 CC InterPro; IPR005481; CPase_L_N.
 CC InterPro; IPR001317; CPS_GATase.
 CC InterPro; IPR002474; CPase_sm_chain.
 CC InterPro; IPR000991; GATase_L.
 CC InterPro; IPR004362; MGS_like.
 CC Pfam; PF00289; CPase_L_chain; 2.
 CC Pfam; PF02786; CPase_L_D2; 2.
 CC Pfam; PF02787; CPase_L_D3; 1.
 CC Pfam; PF00988; CPase_sm_chain; 1.
 CC Pfam; PF00117; GATase; 1.
 CC Pfam; PF02143; MGS; 1.
 CC PRINTS; PR00098; CPSASE.
 CC PRINTS; PR00099; CPSGATASE.
 CC PRINTS; PR00096; GATASE.
 CC TIGRFAMs; TIGR01369; CPSaseII_lrg; 1.
 CC TIGRFAMs; TIGR01368; CPSaseIISmall; 1.
 CC PROSITE; PS00866; CPSASE_1; 2.
 CC PROSITE; PS00867; CPSASE_2; 2.
 CC Ligase; Allosteric enzyme; Repeat; Transit peptide; Mitochondrion;
 CC ATP-binding; Urea cycle; Polymorphism; Disease mutation.
 CC TRANSIT 1 38 MITOCHONDRION (BY SIMILARITY).
 CC CHAIN 39 1500 CARBAMOYL-PHOSPHATE SYNTHASE [AMMONIA].
 CC FT CHAIN 39 1500 ATP (2).
 CC FT NP_BIND 571 626 ATP (2).
 CC FT NP_BIND 718 768 ATP (1).
 CC FT NP_BIND 1113 1171 ATP (2).
 CC FT NP_BIND 1259 1302 ATP (1).
 CC FT DOMAIN 39 219 ANTHRANILATE PHOSPHORIBOSYLTRANSFERASE
 CC FT HOMOLOG.
 CC FT DOMAIN 220 410 GLUTAMINE AMIDOTRANSFERASE-LIKE.
 CC FT DOMAIN 411 1500 CARBAMOYL-PHOSPHATE SYNTHETASE LARGE
 CC FT CHAIN.
 CC FT REPEAT 419 876
 CC FT REPEAT 970 1410
 CC FT VARIANT 337 337
 CC FT VARIANT 344 344
 CC FT VARIANT 344 344
 CC FT VARIANT 544 544
 CC FT VARIANT 544 544
 CC FT CONFLICT 111 111 A -> S (IN REF. 1).
 CC FT CONFLICT 279 279 R -> Q (IN REF. 1).
 CC FT CONFLICT 338 338 G -> C (IN REF. 1).
 CC FT CONFLICT 718 722 RLSRS -> KMSPN (IN REF. 1).

FT CONFLICT 729 729 A -> T (IN REF. 1).
 FT CONFLICT 749 749 E -> G (IN REF. 1).
 FT CONFLICT 1161 1162 EH -> AT (IN REF. 1).
 FT CONFLICT 1204 1205 GD -> EN (IN REF. 1).
 FT CONFLICT 1254 1254 I -> N (IN REF. 1).
 FT CONFLICT 1266 1266 F -> S (IN REF. 1).
 FT CONFLICT 1283 1283 M -> L (IN REF. 1).
 FT CONFLICT 1303 1303 A -> V (IN REF. 1).
 FT CONFLICT 1406 1406 T -> N (IN REF. 1).
 SQ SEQUENCE 1500 AA; 164939 MW; E53R22D77563961D CRC64;
 Query Match 97.8%; Score 266; DB 1; Length 1500;
 Best Local Similarity 98.0%; Pred. No. 6.5e-25;
 Matches 50; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 QLNHEGKLFATEATSDWLNANNVPANVPVAVPQEGNPSLSIRKLRDG 51
 Db 1380 QLNHEGKLFATEATSDWLNANNVPANVPVAVPQEGNPSLSIRKLRDG 1430
 RESULT 2
 CPSM_RAT STANDARD; PRT; 1500 AA.
 ID CPSM_RAT
 AC P07756;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Carbamoyl-phosphate synthase [ammonia], mitochondrial precursor
 DE (EC 6.3.4.16) (Carbamoyl-phosphate synthetase I) (CPSASE I).
 GN CPS1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85261323; PubMed=2991241;
 RA Nyunoya H., Broglie K.E., Widgren E.E., Lusty C.J.;
 RT "Characterization and derivation of the gene coding for mitochondrial
 FT carbamyl phosphate synthetase I of rat.";
 RL J. Biol. Chem. 262:10415-10418(1987).
 RN [2]
 RP SEQUENCE OF 1-42 FROM N.A.
 RX MEDLINE=87280088; PubMed=3038878;
 RA Lagace M., Howell B.W., Burak R., Lusty C.J., Shore G.C.;
 RT "Rat carbamyl-phosphate synthetase I gene. Promoter sequence and
 FT tissue-specific transcriptional regulation in vitro.";
 RL J. Biol. Chem. 262:10415-10418(1987).
 CC -!- FUNCTION: INVOLVED IN THE UREA CYCLE OF UREOTELIC ANIMALS WHERE
 CC THE ENZYME PLAYS AN IMPORTANT ROLE IN REMOVING EXCESS AMMONIA
 CC FROM THE CELL.
 CC -!- CATALYTIC ACTIVITY: 2 ATP + NH(3) + CO(2) + H(2)O = 2 ADP +
 CC phosphate + carbamoyl phosphate.
 CC -!- ENZYME REGULATION: REQUIRES N-ACETYLGLUTAMATE AS AN ALLOSTERIC
 CC ACTIVATOR.
 CC -!- SUBCELLULAR LOCATION: Mitochondrial.
 CC -!- TISSUE SPECIFICITY: PRIMARILY IN THE LIVER AND SMALL INTESTINE.
 CC -!- PTM: 50% OF THE MATURE PROTEIN THAT WAS ISOLATED HAD LEU 39 AS ITS
 CC N-TERMINAL RESIDUE AND 50% HAD SER 40 SUGGESTING TWO ADJACENT
 CC PROCESSING SITES. HOWEVER, THE POSSIBILITY OF PROTEOLYTIC REMOVAL
 CC OF LEU 39 DURING THE ISOLATION OF THE ENZYME CANNOT BE EXCLUDED.
 CC -!- SIMILARITY: TO OTHER CARBAMOYL-PHOSPHATE SYNTHETASES. ALSO
 CC CONTAINS A GLUTAMINE AMIDOTRANSFERASE-LIKE DOMAIN BUT THE CYSTEINE
 CC RESIDUE WHICH IS ESSENTIAL FOR AMIDO TRANSFERASE ACTIVITY, IS
 CC MISSING IN THIS PROTEIN SO THIS DOMAIN IS NO LONGER ACTIVE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2003, 02:03:24 ; Search time 14 seconds
(without alignments)
171.312 Million cell updates/sec

Title: US-09-585-077c-2_COPY_1380_1430

Perfect score: 272

Sequence: 1 OLHNEGFKLFATEATSDWLN.....PSQEQNPSSLIRKLINDG 51

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	266	97.8	1500	1	CPSM_HUMAN
2	266	97.8	1500	1	P07756 rattus norv
3	179.5	66.0	1496	1	CPSM_RANCA
4	80	29.4	618	1	CAB2 METJA
5	71.5	26.3	1115	1	CABE MYCTU
6	69.5	25.6	1071	1	CARB_BACSU
7	69.5	25.6	2185	1	PYRL_DICDI
8	65	23.9	2214	1	PYRL_YEAST
9	64	23.5	314	1	APAH_XANAC
10	64	23.5	452	1	PUR9 THEMA
11	64	23.5	1102	1	CARB_STRCO
12	64	23.5	2244	1	PYRL_SCHPO
13	60.5	22.2	1228	1	SLAP_BACST
14	59.5	21.9	199	1	PEXH_YEAST
15	58.5	21.5	1121	1	CARB_MYCLE
16	58	21.3	2242	1	PYRL_SOUCI
17	56.5	20.8	457	1	BAGA_MOUSE
18	56.5	20.8	811	1	RYFL_YEAST
19	56	20.6	521	1	YF92_MYCPN
20	56	20.6	591	1	ALUH_HUMAN
21	56	20.6	772	1	LMBT_HUMAN
22	56	20.6	1113	1	TCMP_CORGL
23	55.5	20.4	270	1	TCMP_STRGA
24	55	20.2	767	1	HYPE_SYNY3
25	55	20.2	786	1	BTKL_DROME
26	55	20.2	790	1	SNY2_YEAST
27	54.5	20.0	662	1	YFL8_YEAST
28	54.5	20.0	1034	1	BGAL_KLEPN
29	54.5	20.0	1092	1	CARB_ZYMO
30	54.5	20.0	2225	1	PYRL_HUMAN
31	54.5	20.0	2225	1	PYRL_MESAU
32	54	19.9	441	1	PAC1_MOUSE
33	54	19.9	441	1	PAC1_RAT

RESULT 1

ID	CPSM_HUMAN	STANDARD;	PRT;	1500 AA.
AC	P31327; O43774;			
DT	01-JUL-1993 (Rel. 26, Created)			
DT	15-DEC-1998 (Rel. 37, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Carbamoyl-phosphate synthase [ammonial, mitochondrial precursor			
DE	(EC 6.3.4.16) (Carbamoyl-phosphate synthetase I) (CPSase I).			
GN	CPS1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RN	SEQUENCE FROM N.A.			
RC	TISSUE=Liver;			
RX	MEDLINE=92084128; PubMed=1840546;			
RA	Haraguchi Y., Uchino T., Takiguchi M., Endo F., Mori M.,			
RA	Matsuda I.;			
RT	"Cloning and sequence of a cDNA encoding human carbamyl phosphate			
RT	synthetase I: molecular analysis of hyperammonemia.";			
RL	Gene 107:335-340(1991).			
RN	[2]			
RN	SEQUENCE FROM N.A., VARIANT CPS1 DEFICIENCY MET-544, AND VARIANT			
RP	ALA-344.			
RC	TISSUE=Liver;			
RX	MEDLINE=98375696; PubMed=9711878;			
RA	Finckh U., Kohlschuetter A., Schaefer H., Sperhake K., Colombo J.-P.,			
RA	Gal A.;			
RT	"Prenatal diagnosis of carbamoyl phosphate synthetase I deficiency by			
RT	identification of a missense mutation in CPS1.";			
RL	Hum. Mutat. 12:206-211(1998).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RA	Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.			
RL	[4]			
RN	VARIANT CPS1 DEFICIENCY ARG-337.			
RX	MEDLINE=21367742; PubMed=11474210;			
RA	Aoshima T., Kajiya M., Sekido Y., Kikuchi S., Yasuda I., Saheki T.,			
RA	Watanabe K., Shimokata K., Niwa T.;			
RT	"Novel mutations (H337R and 238-362del) in the CPS1 gene cause			
RT	carbamoyl phosphate synthetase I deficiency.";			
RL	Hum. Hered. 52:99-101(2001).			
CC	!- FUNCTION: INVOLVED IN THE UREA CYCLE OF UREOTELIC ANIMALS WHERE			
CC	THE ENZYME PLAYS AN IMPORTANT ROLE IN REMOVING EXCESS AMMONIA			
CC	FROM THE CELL.			
CC	!- CATALYTIC ACTIVITY: 2 ATP + NH(3) + CO(2) + H(2)O = 2 ADP +			
CC	phosphate + carbamoyl phosphate.			
CC	!- ENZYME REGULATION: REQUIRES N-ACETYLGLUTAMATE AS AN ALLOSTERIC			
CC	ACTIVATOR.			
CC	!- SUBCELLULAR LOCATION: Mitochondrial.			
CC	!- TISSUE SPECIFICITY: PRIMARILY IN THE LIVER AND SMALL INTESTINE.			
CC	!- DISEASE: Defects in CPS1 are the cause of an autosomal recessive			
CC	metabolic disorder that cause a type of hyperammonemia. Clinical			

Search completed: September 13, 2003, 02:16:42
Job time : 28.5 secs

[illegible]

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbiag, K.;
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A>Title: Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pa
A|Reference number: AB2950; PMID:20437337; PMID:10984043
A|Accession: B83605
A|Status: preliminary
A|Molecule type: DNA
A|Residues: 1-344 <SFO>
A|Cross-references: GB:AE004470; GB:AE004091; NID:g9946164; PIDN:AAG03710.1; GSPDB:GN
C|Experimental source: strain PAOI
C|Genetics:
A|Gene: PAO321

Query Match 24.8%; Score 67.5; DB 2; Length 344;
Best Local Similarity 37.1%; Pred.No.1;
Matches 13; Conservative 7; Mismatches 14; Indels 1; Gaps 1;

QY 1 QLNHEGFKLFATEATSDWL-NANNVPANVPWPQSQ 34
:::|||| | |||| : ||||::
Db 61 RVHSEGFVFLQNAWDWLATGRSHDMLPIAWPTR 95

RESULT 12
QZBYU2
pyrimidine synthesis protein URA2 - yeast (*Saccharomyces cerevisiae*)
N|Alternate names: protein J0686; protein YJL130C
N|Contents: aspartate carbamoyltransferase (EC 2.1.1.3.2); carbamoyl-phosphate synthetase
C|Species: *Saccharomyces cerevisiae*
C|Date: 30-Sep-1991 #sequence_revision 08-Sep-1995 #text_change 12-Nov-1999
C|Accession: S56912; S05767; S05859; S53182; S13338; S71667; S22790; A33820;
R|Czapluch, C.; Kordes, E.; PuJol, A.; Jauniaux, J.C.
submitted to the Protein Sequence Database, September 1995
A|Reference number: S56891
A|Accession: S56911
A|Molecule type: DNA
A|Residues: 1-276 <CZI>
A|Cross-references: EMBL:Z49405; GSPDB:GN00010; MIPS:YJL130C
R|Katsoulou, C.; Tzermlia, M.; Alexandraki, D.
submitted to the Protein Sequence Database, September 1995
A|Reference number: S56912
A|Accession: S56912
A|Molecule type: DNA
A|Residues: 175-2214 <XAT>
A|Cross-references: EMBL:Z49405; GSPDB:GN00010; MIPS:YJL130C
R|Souciet, J.-L.; Nagt, M.; le Gouar, M.; Lacroute, F.; Potier, S.
Gene 79, 59-70, 1989
A|Title: Organization of the yeast URA2 gene: identification of a defective dihydrodro
A|Reference number: S05766; PMID:89378778; PMID:2570735
A|Accession: S05767
A|Molecule type: DNA
A|Residues: 1-122, A', 124-249, RI', 250, 'SCSMD', 258-269, Y', 271-312, VQ', 315-371, RF',
A', 1596-2214 <SOUI>
A|Cross-references: EMBL:M27174
R|Souciet, J.-L.; Potier, S.; Hubert, J.-C.; Lacroute, F.
Mol. Gen. Genet. 207, 314-319, 1987
A|Title: Nucleotide sequence of the pyrimidine specific carbamoyl phosphate synthetas
A|Reference number: S05859; MOID:87286375; PMID:3039294
A|Accession: S05859
A|Molecule type: DNA
A|Residues: 1-85, D', 87-122, A', 124-249, RI', 250, 'SCSMD', 258-269, Y', 271-312, VQ', 315-371,
506 <SOUI2>
A|Cross-references: EMBL:X05553; NID:g4760; PIDN:CAA29068.1; PID:g4761
A|Note: the authors translated the codon GAC for residue 85 as His
R|Katsoulou, C.; Tzermlia, M.; Alexandraki, D.
submitted to the EMBL Data Library, May 1995
A|Description: The complete sequence of a 40.7 kb segment located on the left arm of
yeast hypothetical proteins.
A|Reference number: S55159
A|Accession: S55182
A|Molecule type: DNA
A|Residues: 175-2214 <KA2>
A|Cross-references: EMBL:X87371; NID:g854542; PIDN:CAA60825.1; PID:g854566

F; 9-1089/Domain: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain hom

```
Query Match          26.3%; Score 71.5; DB 2; Length 1115;
Best Local Similarity 35.2%; Pred. No. 1.3;
Matches 19; Conservative 14; Mismatches 16; Indels 5; Gaps 3;

QY 1 QLNHEGFKLPATEATSDWLNANNVPANPVA---WFSQEQGNPSLSSIRKLIRDG 51
   : | | | : | | | : | | | : | | | : | | | : | | | : | | |
Db 992 RLADQFRVLATEGTAEMRLRNGIPCDQDKRHFEPAPGR-PTMSAV-DAIRAG 1043

RESULT 8
S23738
pyr1-3 protein - slime mold (Dictyostelium discoideum) (fragment)
N:Contains: carbamoyl-phosphate synthase (glutamine-hydrolyzing) (EC 6.3.5.5)
C:Species: Dictyostelium discoideum
C/Date: 08-Jun-1994 #sequence_revision 08-Sep-1995 #text_change 09-Jun-2000
C/Accession: S23738
R/Elgar, G.; Schofield, J.P.
DNA Seq. 2, 219-226, 1992
A/Title: Carbamoyl-phosphate synthetase (CPSase) in the PYR1-3 multigene of Dictyoste
A:Reference number: S23738; MUID:92329976; PMID:1627825
A/Accession: S23738
A/Molecule type: DNA
A/Residues: 1-1042 <BLG>
A/Cross-references: EMBL:X55433; NID:g7244; PIDN:CAA39077.1; PID:g7245
C:Superfamily: rudimentary enzyme; aspartate/ornithine carbamoyltransferase homology;
arabamoyl-phosphate synthase (glutamine-hydrolyzing) large chain homology; carbamoyl-p
C/Keywords: ligase
F:1-1042/Domain: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain hom
F:1-1042/Domain: carbamoyl-phosphate synthase (ammonia) homology <CPA>
F:1-453/Domain: biotin carboxylase homology <BC1>
F:543-990/Domain: biotin carboxylase homology <BC2>

Query Watch          25.6%; Score 69.5; DB 2; Length 1042;
Best Local Similarity 35.3%; Pred. No. 2.2;
Matches 12; Conservative 9; Mismatches 12; Indels 1; Gaps 1;

QY 6 GFKLFATEATSDWLNANNVPANPVAWFSQEQ-GQN 38
   : | | | : | | | : | | | : | | | : | | | : | | |
Db 950 GYTFEGTGTADEFYSENGVPYTLNWDEEDLGEN 983

RESULT 9
F39845
carbamoyl-phosphate synthase (glutamine-hydrolyzing) (EC 6.3.5.5), pyrimidine-repress
N/Alternate names: carbamoyl-phosphate synthetase (catalytic chain) PYRAB
C/Species: Bacillus subtilis
C/Date: 20-Mar-1992 #sequence_revision 20-Mar-1992 #text_change 02-Sep-2000
C/Accession: F39845; A69686
R/Quinn, C.L.; Stephenson, B.T.; Switzer, R.L.
J. Biol. Chem. 265, 9113-9127, 1991
A/Title: Functional organization and nucleotide sequence of the Bacillus subtilis pyr
A:Reference number: A39845; MUID:91225016; PMID:1709162
A/Accession: F39845
A/Molecule type: DNA
A/Residues: 1-1071 <GUI>
A/Cross-references: GB:W59757; NID:g4887706; PIDN:AAA21270.1; PID:g143390
R/Kunst, F.; Ogawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari
Nature 390, 249-256, 1997
A:Authors: Boulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.;
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
y, M.; Odawa, K.; Ogigawa, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
t.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Zumbstein, E.; Yoshikawa, H.; Zumbstein, E.
A/Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
```

QY 2 LHNEGFKLFAATSDWLNANNVPANVPWPSEQNPSLSSTKLIIRGG 51
I :|||||:|||||:| | |||||:| | |||:|
Db 1378 LKEGFKLATATADNLANDITAPVWPSEQGS-GPSYIKLIKEG 1426

RESULT 4
S53602
carbamoyl-phosphate synthase (glutamine-hydrolyzing) (EC 6.3.5.5) III precursor, hepa
N/Alternate names: carbamyl phosphate synthetase III
C/Species: Squalus acanthias (spiny dogfish)
C/Date: 28-Oct-1996 #sequence_revision 07-Feb-1997 #text_change 07-Dec-1999
C/Acession: S53602; S72160; S71062
R/Hong, J.; Salo, W.L.; Lusty, C.J.; Anderson, P.M.
J. Mol. Biol. 243, 131-140, 1994
A:/Title: Carbamyl phosphate synthetase III, an evolutionary intermediate in the trans
A:/Reference number: S53602; MUID:95018247; PMID:7932737
A/Acession: S53602
A:/Status: nucleic acid sequence not shown
A:/Molecule type: mRNA
A:/Residues: 1-1502 <HOM1>
A:/Cross-references: EMBL:L31362
A/Acession: S72160
A:/Molecule type: protein
A:/Residues: 39-47 <HON2>
R/Anderson, P.M.
submitted to the EMBL Data Library, September 1994
A:/Reference number: S71062
A/Acession: S71062
A:/Molecule type: mRNA
A:/Residues: 1-888, E*, 890-1502 <AND>
A:/Cross-references: EMBL:L31362; NID:g1256714; PIDN:AAA96435.1; PID:g530209
C/Genetics:
A:/Genome: nuclear
C/Superfamily: carbamoyl-phosphate synthase (ammonia); biotin carboxylase homology; c
omology; carbamoyl-phosphate synthase (glutamine-hydrolyzing) small chain homology; t
C/Keywords: arginine biosynthesis; ligase; mitochondrion; pyrimidine nucleotide biosy
E:1-39/Domain: transit peptide (mitochondrion) #status predicted <TNP>
F:39-1502/Product: carbamyl-phosphate synthetase (glutamine-hydrolysing) III #status
F:48-1480/Domain: carbamoyl-phosphate synthase (ammonia) homology <CPA>
F:48-396/Domain: carbamoyl-phosphate synthase (glutamine-hydrolyzing) small chain hom
F:220-396/Domain: trps homology <TRG>
F:425-1477/Domain: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain h
F:425-887/Domain: biotin carboxylase homology <BCL>
F:976-1427/Domain: biotin carboxylase homology <BC2>
F:294/Active site: Cys #status predicted

Query Match 59.9%; Score 163; DB 2; Length 1502;
Best Local Similarity 56.9%; Pred. No. 6e-12;
Matches 29; Conservative 7; Mismatches 15; Indels 0; Gaps 0;

QY 1 QLHNEGFKLFATEANSDWLNANNVPANVPWPSEQNPSLSSTKLIIRGG 51
||| :|||:||||| |||||:| | |||||:| | ||| |||||
Db 1381 QLHFQGYKLYATEGTSLNDVNVDTPAVSNPTAEHDSSASPFSTKIHDG 1431

RESULT 5
D64472
carbamoyl-phosphate synthase, pyrimidine-specific, large subunit (EC 6.3.-.-) - Metha
C/Species: Methanococcus jannaschii
C/Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C/Acession: D64472
R/Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak
; Raich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,
rson, J.D.; Sadov, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A/Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese
A:/Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc
A:/Reference number: A64300; MUID:96337999; PMID:8688087
A/Acession: D64472
A:/Status: preliminary; nucleic acid sequence not shown; translation not shown
A:/Molecule type: DNA
A:/Residues: 1-618 <BLU>
A:/Cross-references: GB:D67578; GB:L77117; NID:g2826403; PIDN:AAB99391.1; PID:g1592073

Search completed: September 13, 2003, 02:12:44
Job time : 55 secs


```

RESULT 11
ABP39849      Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4694.
ID            ABP39849 standard; Protein: 228 AA.
XX
AC             ABP39849;
XX
DT              24-JUL-2002   (first entry)
XX
DE               Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4694.
XX
KW                Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
KN                 antibacterial; gene therapy.
XX
OS                  Staphylococcus epidermidis.
XX
US        US6380370-B1..
PN
XX
PD          30-APR-2002.
XX
PF           13-AUG-1998;    98US-0134001.
XX
PR           14-AUG-1997;    97US-055779P.
PP           08-NOV-1997;    97US-064964P.
XX
PA         (GENO-) GENOME THERAPEUTICS CORP.
XX
PI       Doucette-Stamm LA, Bush D;
XX
DR     WPI; 2002-381255/41.
DR     N-PSTDB; ABN92394.
XX
PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis
PE polypeptide, useful for diagnosing and treating bacterial infections -
XX
FS Disclosure; SEQ ID 4694; 267pp; English.
XX
CC
XX
ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
frame (ORF) nucleic acid sequences which encode the amino acid sequences
given in ABP35124 to ABP37960. The S. epidermidis sequences have
antibacterial activity and can be used in gene therapy. The sequences
can also be used in the diagnosis and treatment of bacterial infections,
particularly S. epidermidis infections. The sequences can be used to
screen for compounds able to interfere with the S. epidermidis life
cycle or inhibit S. epidermidis infection.
CC
N.B. The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from the
USPTO web site.
XX
SQ Sequence 228 AA;
XX
Query Match                      23.2%; Score 63; DB 23; Length 228;
Best Local Similarity            26.0%; Pred.No. 3.5;
Matches 13; Conservative 11; Mismatches 22; Indels 4; Gaps 1
XX
QY 1 QLNHEGFKLFATTSOWLNNANVPNVAMPSPGOEGNSLSIRKLIRD 50
| | | : | | | | | : | : | : | : | | | | : |
Db 157 KLRKEDFDILGTENVINWLTKSIKSIOIS-----EGTNVEIKTELLKD 202
XX
RESULT 12
ABP40650
ID            ABP40650 standard; Protein: 228 AA.
XX
AC             ABP40650;
XX
DT              24-JUL-2002   (first entry)
XX
DE               Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5495.
XX

```


CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 344 AA;

Query Match 24.8%; Score 67.5; DB 22; Length 344;
 Best Local Similarity 37.1%; Pred. No. 1.5;
 Matches 13; Conservative 7; Mismatches 14; Indels 1; Gaps 1;

QY 1 QLNHEGFKLPATEATSDWL-NANNYPANVPANVPSQ 34
 Db 61 RVHSEGFVFTQNAQWDNLATGRSHDMLPIAWPTR 95

RESULT 7

AA013747
 ID AA013747 standard; Protein; 43 AA.

XX AA013747;

XX 06-NOV-2001 (first entry)

DE Human polypeptide SEQ ID NO 27639.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation.

OS Homo sapiens.

XX WO200164835-A2.

XX 07-SEP-2001.

XX 26-FEB-2001; 2001WO-US04927.

XX 28-FEB-2000; 2000US-0515126.

XX 18-MAY-2000; 2000US-0577409.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-514838/56.

XX N-PSDB; AA193678.

PT Isolated nucleic acids and polypeptides, useful for preventing
 PT diagnosing and treating e.g. leukaemia, inflammation and immune
 PT disorders -

PS Claim 20; SEQ ID NO 27639; 1399pp + Sequence Listing; English.

XX The invention relates to human polynucleotides (AA179941-AA193841) and
 CC the encoded proteins (AA000010-AA013910) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and

CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 43 AA;

Query Match 24.3%; Score 56; DB 22; Length 43;

Best Local Similarity 34.9%; Pred. No. 0.16;

Matches 15; Conservative 6; Mismatches 18; Indels 4; Gaps 1;

QY 1 QLNHEGFKLPATEATSDWL-NANNYPANVPANVPSQEQNDPSLSS 43

Db 5 QIHKLFKPRILGWITWL---TPVIPAPWAEERGRSPSS 43

RESULT 8

AA009946
 ID AA009946 standard; Protein; 94 AA.

XX AA009946;

XX 06-NOV-2001 (first entry)

DE Human polypeptide SEQ ID NO 23838.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation.

OS Homo sapiens.

XX WO200164835-A2.

XX 07-SEP-2001.

XX 26-FEB-2001; 2001WO-US04927.

XX 28-FEB-2000; 2000US-0515126.

XX 18-MAY-2000; 2000US-0577409.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-514838/56.

XX N-PSDB; AA189877.

PT Isolated nucleic acids and polypeptides, useful for preventing
 PT diagnosing and treating e.g. leukaemia, inflammation and immune
 PT disorders -

PS Claim 20; SEQ ID NO 23838; 1399pp + Sequence Listing; English.

XX The invention relates to human polynucleotides (AA179941-AA193841) and
 CC the encoded proteins (AA000010-AA013910) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 94 AA;

Query Match 23.5%; Score 64; DB 22; Length 94;

PT CDNA contg. total human cDNA sequence
 XX Disclosure; Fig 2-9; 12pp; Japanese.
 XX
 CC This sequence is the human carbamyl phosphate synthetase I (CPSI)
 CC protein. The DNA encoding this peptide or fragments of it may be
 CC used as probes to detect mutations in the CPSI gene. CPSI deficiency
 CC disease may be detected by using overlapping cDNA representing the
 CC full length cDNA sequence of human CPSI. The DNA sequence was
 CC isolated using three amplified fragments derived from protein coding
 CC regions of the rat CPSI cDNA as probes.
 XX
 XX Sequence 1500 AA;
 SQ
 Query Match 100.0%; Score 272; DB 14; Length 1500;
 Best Local Similarity 100.0%; Pred. No. 1.5e-27;
 Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 QLNHEGKLFATEATSDWLNANNYPANVPVAPVPSQEQNPSSLSIRKLIRDG 51
 DB 1380 QLNHEGKLFATEATSDWLNANNYPANVPVAPVPSQEQNPSSLSIRKLIRDG 1430
 RESULT 2
 AAB49222
 ID AAB49222 standard; protein; 1500 AA.
 XX
 AC AAB49222;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Human N1405 carbamyl phosphate synthase I protein.
 XX
 KW Urea cycle; carbamyl phosphate synthase I; CPSI; hepatitis;
 KW sclerosis; pulmonary hypertension; bone marrow; gene therapy.
 XX
 OS Homo sapiens.
 OS
 PN WO200073322-A1.
 XX
 PD 07-DEC-2000.
 XX
 PF 01-JUN-2000; 2000WO-US15079.
 XX
 PR 01-JUN-1999; 99US-0323472.
 XX
 PA (UYVA-) UNIV VANDERBILT.
 XX
 PI Summar ML, Christman BW;
 XX
 DR WPI; 2001-049926/06.
 XX
 XX Detecting susceptibility to suboptimal urea cycle function, e.g. bone
 PT marrow transplant toxicity, comprises identifying a polymorphism in the
 PT gene for carbamylphosphate synthase
 XX
 PS Claim 57; Fig 12; 171pp; English.
 XX
 XX The present invention relates to screening for susceptibility to
 CC suboptimal urea cycle function or to bone marrow transplant toxicity
 CC by detecting a polymorphism in the carbamyl phosphate synthase I
 CC (CPSI) gene. The method is used to detect subjects at risk of
 CC hepatitis, sclerosis, pulmonary hypertension and bone marrow
 CC transplant toxicity. These conditions can be treated or prevented
 CC by administration of a nitric oxide precursor or by gene
 CC therapy (administration of sequences that encode CPSI).
 XX
 XX Sequence 1500 AA;
 SQ
 Query Match 100.0%; Score 272; DB 22; Length 1500;
 Best Local Similarity 100.0%; Pred. No. 1.5e-27;
 Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 QLNHEGKLFATEATSDWLNANNYPANVPVAPVPSQEQNPSSLSIRKLIRDG 51
 DB 1380 QLNHEGKLFATEATSDWLNANNYPANVPVAPVPSQEQNPSSLSIRKLIRDG 1430
 RESULT 3
 AAB49224
 ID AAB49224 standard; protein; 1500 AA.
 XX
 AC AAB49224;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Human carbamyl phosphate synthase protein.
 XX
 KW Urea cycle; carbamyl phosphate synthase I; CPSI; hepatitis;
 KW sclerosis; pulmonary hypertension; bone marrow; gene therapy.
 XX
 OS Homo sapiens.
 OS
 PN WO200073322-A1.
 XX
 PD 07-DEC-2000.
 XX
 PF 01-JUN-2000; 2000WO-US15079.
 XX
 PR 01-JUN-1999; 99US-0323472.
 XX
 PA (UYVA-) UNIV VANDERBILT.
 XX
 PI Summar ML, Christman BW;
 XX
 DR WPI; 2001-049926/06.
 XX
 XX Detecting susceptibility to suboptimal urea cycle function, e.g. bone
 PT marrow transplant toxicity, comprises identifying a polymorphism in the
 PT gene for carbamylphosphate synthase
 XX
 PS Claim 57; Page 152-155; 171pp; English.
 XX
 XX The present invention relates to screening for susceptibility to
 CC suboptimal urea cycle function or to bone marrow transplant toxicity
 CC by detecting a polymorphism in the carbamyl phosphate synthase I
 CC (CPSI) gene. The method is used to detect subjects at risk of
 CC hepatitis, sclerosis, pulmonary hypertension and bone marrow
 CC transplant toxicity. These conditions can be treated or prevented
 CC by administration of a nitric oxide precursor or by gene
 CC therapy (administration of sequences that encode CPSI).
 XX
 XX Sequence 1500 AA;
 SQ
 Query Match 100.0%; Score 272; DB 22; Length 1500;
 Best Local Similarity 100.0%; Pred. No. 1.5e-27;
 Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 QLNHEGKLFATEATSDWLNANNYPANVPVAPVPSQEQNPSSLSIRKLIRDG 51
 DB 1380 QLNHEGKLFATEATSDWLNANNYPANVPVAPVPSQEQNPSSLSIRKLIRDG 1430
 RESULT 4
 AAB49225
 ID AAB49225 standard; protein; 1500 AA.
 XX
 AC AAB49225;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Human carbamyl phosphate synthase protein.
 XX
 KW Urea cycle; carbamyl phosphate synthase I; CPSI; hepatitis;
 KW sclerosis; pulmonary hypertension; bone marrow; gene therapy.
 XX
 OS Homo sapiens.
 OS
 PN WO200073322-A1.
 XX
 PD 07-DEC-2000.
 XX
 PF 01-JUN-2000; 2000WO-US15079.
 XX
 PR 01-JUN-1999; 99US-0323472.
 XX
 PA (UYVA-) UNIV VANDERBILT.
 XX
 PI Summar ML, Christman BW;
 XX
 DR WPI; 2001-049926/06.
 XX
 XX Detecting susceptibility to suboptimal urea cycle function, e.g. bone
 PT marrow transplant toxicity, comprises identifying a polymorphism in the
 PT gene for carbamylphosphate synthase
 XX
 PS Claim 57; Fig 12; 171pp; English.
 XX
 XX The present invention relates to screening for susceptibility to
 CC suboptimal urea cycle function or to bone marrow transplant toxicity
 CC by detecting a polymorphism in the carbamyl phosphate synthase I
 CC (CPSI) gene. The method is used to detect subjects at risk of
 CC hepatitis, sclerosis, pulmonary hypertension and bone marrow
 CC transplant toxicity. These conditions can be treated or prevented
 CC by administration of a nitric oxide precursor or by gene
 CC therapy (administration of sequences that encode CPSI).
 XX
 XX Sequence 1500 AA;
 SQ
 Query Match 100.0%; Score 272; DB 22; Length 1500;
 Best Local Similarity 100.0%; Pred. No. 1.5e-27;
 Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLNHEGKLFATEATSDWLNANNYPANVPVAPVPSQEQNPSSLSIRKLIRDG 51
 DB 1380 QLNHEGKLFATEATSDWLNANNYPANVPVAPVPSQEQNPSSLSIRKLIRDG 1430
 RESULT 3
 AAB49224
 ID AAB49224 standard; protein; 1500 AA.
 XX
 AC AAB49224;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Human carbamyl phosphate synthase protein.
 XX
 KW Urea cycle; carbamyl phosphate synthase I; CPSI; hepatitis;
 KW sclerosis; pulmonary hypertension; bone marrow; gene therapy.
 XX
 OS Homo sapiens.
 OS
 PN WO200073322-A1.
 XX
 PD 07-DEC-2000.
 XX
 PF 01-JUN-2000; 2000WO-US15079.
 XX
 PR 01-JUN-1999; 99US-0323472.
 XX
 PA (UYVA-) UNIV VANDERBILT.
 XX
 PI Summar ML, Christman BW;
 XX
 DR WPI; 2001-049926/06.
 XX
 XX Detecting susceptibility to suboptimal urea cycle function, e.g. bone
 PT marrow transplant toxicity, comprises identifying a polymorphism in the
 PT gene for carbamylphosphate synthase
 XX
 PS Claim 57; Page 152-155; 171pp; English.
 XX
 XX The present invention relates to screening for susceptibility to
 CC suboptimal urea cycle function or to bone marrow transplant toxicity
 CC by detecting a polymorphism in the carbamyl phosphate synthase I
 CC (CPSI) gene. The method is used to detect subjects at risk of
 CC hepatitis, sclerosis, pulmonary hypertension and bone marrow
 CC transplant toxicity. These conditions can be treated or prevented
 CC by administration of a nitric oxide precursor or by gene
 CC therapy (administration of sequences that encode CPSI).
 XX
 XX Sequence 1500 AA;
 SQ
 Query Match 100.0%; Score 272; DB 22; Length 1500;
 Best Local Similarity 100.0%; Pred. No. 1.5e-27;
 Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 QLNHEGKLFATEATSDWLNANNYPANVPVAPVPSQEQNPSSLSIRKLIRDG 51
 DB 1380 QLNHEGKLFATEATSDWLNANNYPANVPVAPVPSQEQNPSSLSIRKLIRDG 1430
 RESULT 4
 AAB49225
 ID AAB49225 standard; protein; 1500 AA.
 XX
 AC AAB49225;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Human carbamyl phosphate synthase protein.
 XX
 KW Urea cycle; carbamyl phosphate synthase I; CPSI; hepatitis;
 KW sclerosis; pulmonary hypertension; bone marrow; gene therapy.
 XX
 OS Homo sapiens.
 OS
 PN WO200073322-A1.
 XX
 PD 07-DEC-2000.
 XX
 PF 01-JUN-2000; 2000WO-US15079.
 XX
 PR 01-JUN-1999; 99US-0323472.
 XX
 PA (UYVA-) UNIV VANDERBILT.
 XX
 PI Summar ML, Christman BW;
 XX
 DR WPI; 2001-049926/06.
 XX
 XX Detecting susceptibility to suboptimal urea cycle function, e.g. bone
 PT marrow transplant toxicity, comprises identifying a polymorphism in the
 PT gene for carbamylphosphate synthase
 XX
 PS Claim 57; Page 152-155; 171pp; English.
 XX
 XX The present invention relates to screening for susceptibility to
 CC suboptimal urea cycle function or to bone marrow transplant toxicity
 CC by detecting a polymorphism in the carbamyl phosphate synthase I
 CC (CPSI) gene. The method is used to detect subjects at risk of
 CC hepatitis, sclerosis, pulmonary hypertension and bone marrow
 CC transplant toxicity. These conditions can be treated or prevented
 CC by administration of a nitric oxide precursor or by gene
 CC therapy (administration of sequences that encode CPSI).
 XX
 XX Sequence 1500 AA;
 SQ
 Query Match 100.0%; Score 272; DB 22; Length 1500;
 Best Local Similarity 100.0%; Pred. No. 1.5e-27;
 Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2003, 00:29:37 ; Search time 52 Seconds
(without alignments)
155.674 Million cell updates/sec

Title: US-09-585-077C-2_COPY_1380_1430

Perfect score: 272
Sequence: 1 QLNNEGFKLFATEAGSDWLN.....PSQEGQNFSLSIRKLRDG 51

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

Result No.	Score	Query Match	Length	DB ID	Description
1	272	100.0	1500	14	AA30636 hCPSI. Homo sapie
2	272	100.0	1500	22	AA49222 Human N1405 carbam
3	272	100.0	1500	22	AA49224 Human carbamyl pho
4	272	100.0	1500	22	AA49225 Human carbamyl pho
5	266	97.8	1500	22	AA49223 Human T1405 carbam
6	67.5	24.8	344	22	AA33563 Pseudomonas aerugi
7	66	24.3	43	22	AA013747 Human polypeptide
8	64	23.5	94	22	AA009946 Human polypeptide
9	63.5	23.3	692	22	ABG09932 Novel human diagno

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

10	63.5	23.3	1798	22	ABG09933	Novel human diagno
11	63	23.2	228	23	ABP39849	Staphylococcus epi
12	63	23.2	228	23	ABP40650	Staphylococcus epi
13	63	23.2	619	23	AAO22174	Ramoplanin biosynt
14	61	22.4	60	22	AAO11745	Human polypeptide
15	61	22.4	3095	23	AAE20788	Rat C3b/C4b comple
16	60.5	22.2	1228	17	AA77673	S-layer protein en
17	60.5	22.2	1228	18	AAW22862	Bacillus stearothe
18	60.5	22.2	1228	20	AAW32252	B. stearothermophi
19	60.5	22.2	1228	21	AAAB10625	B. stearothermophi
20	60.5	22.2	1698	20	AA31381	HEV-US1 ORF1 prote
21	60	22.1	59	22	AA82976	Human immune/haema
22	60	22.1	70	22	AAU18578	Human lung antigen
23	60	22.1	70	22	AAU17846	Novel human respir
24	60	22.1	160	22	AAU48056	Propionibacterium
25	60	22.1	340	24	ABU02808	S. pneumoniae type
26	59.5	21.9	94	24	ABP75436	Human secretory po
27	59	21.7	86	22	AAU29913	Novel human secret
28	59	21.7	339	23	AAU03138	Streptococcus pyog
29	59	21.7	339	23	ABP28098	Human myotonic dys
30	59	21.7	690	23	ABP36682	Human protein sequ
31	59	21.7	1020	23	ABP97958	Large subunit of t
32	58	21.3	1018	21	AAU07773	B. lactofermentum
33	58	21.3	1018	22	AA850543	Polyptide with c
34	58	21.3	1113	23	AAU79297	Human protein sequ
35	57.5	21.1	463	22	AA892562	Human normal ovari
36	57.5	21.1	725	20	AA59724	Human protein sequ
37	57.5	21.1	877	22	AA893668	Human apoptosis re
38	57.5	21.1	1181	21	AA82707	Human ORFX ORF281
39	57.5	21.1	1182	21	AA842517	Human apoptosis re
40	57.5	21.1	1220	21	AA842708	Human vesicle traf
41	57.5	21.1	1220	22	AAE04770	Novel human secret
42	57.5	21.1	1232	22	AAU32866	Aspergillus fumiga
43	57.5	21.1	1435	24	ABJ26067	Human secreted pro
44	57	21.0	55	21	AA832094	Human polypeptide
45	57	21.0	84	22	AAO10311	

ALIGNMENTS

RESULT 1	
AA30636	
ID	AA30636 standard; Protein; 1500 AA.
AC	AA30636;
XX	
DT	06-MAY-1993 (first entry)
XX	
DE	hCPSI.
XX	
KW	Human; carbamyl phosphate synthetase I; CPSI; probe; mutation;
KW	CPSI deficiency disease; rat.
XX	
OS	Homo sapiens.
XX	
PN	JP04335889-A.
XX	
PD	24-NOV-1992.
XX	
PF	09-MAY-1991; 91JP-0135902.
XX	
PR	09-MAY-1991; 91JP-0135902.
XX	
PA	(HARA/) HARAGUCHI Y.
PA	(MATS/) MATSUDA I.
XX	(MORI/) MORI M.
DR	WPI; 1993-006237/01.
DR	N-PSDB; AAQ34768.
XX	
PT	Carbamyl phosphate synthetase I gene - used to detect diseases
PT	caused by carbamyl phosphate synthetase by using overlapped

```
; LENGTH: 2193
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-103660

Query Match
  33.8%; Score 27.4; DB 13; Length 2193;
Best Local Similarity 59.7%; Pred. No. 1.9;
Matches 46; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 5 CATCAGACTGGCTCAACGCCAACAAATGTCCTGCCAACCCAGTCGGCATGGCCGCTCTCAAG 64
   ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 341 CATCTGAGCTTTCACCTCCGAAATAGACCCCTCCCATCTTGAGACCTGCAAGSGTCAAG 282
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 65 AAGCAGAGAAATCCCGAGC 81
   ||||| ||||| ||||| |||||
Db 281 AAGGTAAGATCACATC 265

RESULT 14
US-10-207-655-78/c
; Sequence 78, Application US/10207655
; Publication No. US20030118592A1
; GENERAL INFORMATION:
; APPLICANT: Ledbetter, Jeffrey A.
; TITLE OF INVENTION: BINDING DOMAIN-IMMUNOGLOBULIN FUSION PROTEINS
; FILE REFERENCE: 390069.401C1
; CURRENT APPLICATION NUMBER: US/10/207,655
; CURRENT FILING DATE: 2002-07-25
; NUMBER OF SEQ ID NOS: 426
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 78
; LENGTH: 6471
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-207-655-78

Query Match
  33.8%; Score 27.4; DB 14; Length 6471;
Best Local Similarity 59.7%; Pred. No. 2.2;
Matches 46; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 5 CATCAGACTGGCTCAACGCCAACAAATGTCCTGCCAACCCAGTCGGCATGGCCGCTCTCAAG 64
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 332 CATCTGAGCTTTCACCTCCGAAATAGACCCCTCCCATCTTGAGACCTGCAAGSGTCAAG 273
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 65 AAGCAGAGAAATCCCGAGC 81
   ||||| ||||| ||||| |||||
Db 272 AAGGTAAGATCACATC 256

RESULT 15
US-10-068-674-1/c
; Sequence 1, Application US/10068674
; Publication No. US2002017203A1
; GENERAL INFORMATION:
; APPLICANT: Kevirikko, Kari
; APPLICANT: Pihlajaniemi, Taina A.
; TITLE OF INVENTION: 2 SUBUNIT OF PROLYL-4-HYDROXYLASE,
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCE ENCODING SUCH SUBUNIT AND METHODS
; TITLE OF INVENTION: FOR PRODUCING THE SAME
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/068,674
; FILING DATE: 06-FEB-2002
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/633,879
; FILING DATE: 10-MAR-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Halluin, Albert P.
; REGISTRATION NUMBER: 25,227
; REFERENCE/DOCKET NUMBER: 8389-041
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-854-3660
; TELEFAX: 415-854-3694
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2168 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 151..1761
US-10-068-674-1

Query Match
  32.6%; Score 26.4; DB 13; Length 2168;
Best Local Similarity 61.8%; Pred. No. 4.4;
Matches 42; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 9 AGACTGGCTCAACGCCAACAAATGTCCTGCCAACCCAGTCGGCATGGCCGCTCTCAAGAAG 68
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2005 AGTGTGGCTCTACCGCAGCCCATGCCAAGACACCCCCCAGAGACCCCGGCGAGGTGTAGCTGG 1946
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 69 ACAGAATC 76
   ||||| ||||| ||||| |||||
Db 1945 TCAGAGTC 1938
```

Search completed: September 13, 2003, 02:10:57
Job time : 166 secs


```
Query Match      49.6%; Score 40.2; DB 10; Length 242;
Best Local Similarity 93.3%; Pred. No. 2.3e-05;
Matches 42; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 37 GCCACCCAGTGGCGCCCTCTCAGAGGACAGATCCAGC 81
    |||
Db 1 GCTACCCAGTGGCGCCATCTCAAGAGGACAGATCCAGC 45
    |||

RESULT 10
US-09-801-274-1768
; Sequence 1768, Application US/09801274
; Patent No. US20020032319A1
; GENERAL INFORMATION:
; APPLICANT: Cargill, Michele
; APPLICANT: Ireland, James S.
; APPLICANT: Lander, Eric S.
; TITLE OF INVENTION: HUMAN SINGLE NUCLEOTIDE POLYMORPHISMS
; FILE REFERENCE: 2825.2009-001
; CURRENT APPLICATION NUMBER: US/09/801,274
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 60/187,510
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: US 60/206,129
; PRIOR FILING DATE: 2000-05-22
; NUMBER OF SEQ ID NOS: 1902
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1768
; LENGTH: 31
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-801-274-1768

Query Match      37.8%; Score 30.6; DB 9; Length 31;
Best Local Similarity 96.8%; Pred. No. 0.067;
Matches 30; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 26 ACAATGTCCCTGCCAACCCAGTGGCATGGCC 56
    |||
Db 1 ACAATGTCCCTGCCAACCCAGTGGCATGGCC 31
    |||

RESULT 11
US-09-764-891-5507
; Sequence 5507, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5507
; LENGTH: 3591
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-5507

Query Match      36.3%; Score 29.4; DB 11; Length 3591;
Best Local Similarity 63.4%; Pred. No. 0.36;
Matches 45; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 10 GACTGGCTCAACGCCCAACATCTCCCTGCCAACCCAGTGGCATGGCGCTCTCAGAGGA 69
    |||
Db 1746 GACTGGCTGGACCCAGCCAGTCTCTCCGCTCACCCTCTCTCTGGCCCTCACTGGAGGCA 1805
    |||

QY 70 CAGATCCCGAG 80
    |||
Db 1806 GGGACACTCAG 1816
```

```
RESULT 12
US-10-027-632-97735/c
; Sequence 97735, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 97735
; LENGTH: 2193
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-97735

Query Match      33.8%; Score 27.4; DB 13; Length 2193;
Best Local Similarity 59.7%; Pred. No. 1.9;
Matches 46; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 5 CATCAGACTGGCTCAACGCCAAGATGTCCCTGCCAACCCAGTGGCATGGCGCTCTCAAG 64
    |||
Db 341 CATCTGAGTCTTTCACCTCCGATAGCCCTCCATCTTGAGACCTGCAGGTGCAG 282
    |||

QY 65 AAGCAGCAATCCAGC 81
    |||
Db 281 AGAGGTAGATCATC 265
    |||

RESULT 13
US-10-027-632-103660/c
; Sequence 103660, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 103660
```

QY 1 GCCACATCAGACTGGCTCAACGCCCAACAATGTCCTCGCCACACCCAGTGGCATGGCCGTCT 60
|||||
Db 157 GCCACATCAGACTGGCTCAACGCCCAACAATGTCCTCGCCACACCCAGTGGCATGGCCGTCT 216
|||||
QY 61 CAAGAAGGACAGAAATCCCGAGC 81
|||||
Db 217 CAAGAAGGACAGAAATCCCGAGC 237
|||||

RESULT 6

US-10-017-754-93

; Sequence 93, Application US/10017754

; Publication No. US20030054363A1

; GENERAL INFORMATION:

; APPLICANT: Henderson, Robert A.

; APPLICANT: Wang, Tongtong

; APPLICANT: Watanabe, Yoshihiro

; APPLICANT: Johnson, Jeffrey C.

; APPLICANT: Retter, Marc W.

; APPLICANT: Manerakis, Margarita

; APPLICANT: Carter, Darrick

; APPLICANT: Fanger, Gary R.

; APPLICANT: Vedvick, Thomas S.

; APPLICANT: Bangur, Chaitanya S.

; APPLICANT: McNabb, Andria

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER

; FILE REFERENCE: 210121.478C18

; CURRENT APPLICATION NUMBER: US/10/017,754

; CURRENT FILING DATE: 2001-10-29

; NUMBER OF SEQ ID NOS: 2004

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 93

; LENGTH: 531

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-017-754-93

Query Match 98.0%; Score 79.4; DB 14; Length 531;
Best Local Similarity 98.8%; Pred. No. 6.3e-20;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAACGCCCAACAATGTCCTCGCCACACCCAGTGGCATGGCCGTCT 60
|||||
Db 157 GCCACATCAGACTGGCTCAACGCCCAACAATGTCCTCGCCACACCCAGTGGCATGGCCGTCT 216
|||||

QY 61 CAAGAAGGACAGAAATCCCGAGC 81
|||||
Db 217 CAAGAAGGACAGAAATCCCGAGC 237
|||||

RESULT 7

US-10-101-510-553

; Sequence 553, Application US/10101510

; Publication No. US20030148295A1

; GENERAL INFORMATION:

; APPLICANT: WAN, JACKSON

; APPLICANT: WANG, YIXIN

; TITLE OF INVENTION: EXPRESSION PROFILES AND METHODS OF USE

; FILE REFERENCE: 15117.0012

; CURRENT APPLICATION NUMBER: US/10/101,510

; CURRENT FILING DATE: 2002-03-20

; PRIOR APPLICATION NUMBER: 60/276,947

; PRIOR FILING DATE: 2001-03-20

; NUMBER OF SEQ ID NOS: 805

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 553

; LENGTH: 5772

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: modified_base

; LOCATION: (5240)..(5365)

; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-101-510-553

Query Match 98.0%; Score 79.4; DB 12; Length 5772;
Best Local Similarity 98.8%; Pred. No. 8.8e-20;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAACGCCCAACAATGTCCTCGCCACACCCAGTGGCATGGCCGTCT 60
|||||
Db 4310 GCCACATCAGACTGGCTCAACGCCCAACAATGTCCTCGCCACACCCAGTGGCATGGCCGTCT 4369
|||||

QY 61 CAAGAAGGACAGAAATCCCGAGC 81

|||||

Db 4370 CAAGAAGGACAGAAATCCCGAGC 4390

RESULT 8

US-09-960-352-12478

; Sequence 12478, Application US/09960352

; Patent No. US20020137139A1

; GENERAL INFORMATION:

; APPLICANT: Warren, Wesley C.

; APPLICANT: Tao, Nengbing

; APPLICANT: Byatt, John C.

; APPLICANT: Mathialagan, Nagappan

; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION

; FILE REFERENCE: 16511.006/37-21(10298)C

; CURRENT APPLICATION NUMBER: US/09/960,352

; CURRENT FILING DATE: 2001-09-24

; NUMBER OF SEQ ID NOS: 15112

; SEQ ID NO 12478

; LENGTH: 434

; TYPE: DNA

; ORGANISM: Bos taurus

; OTHER INFORMATION: Clone ID: 53-LIB34-037-Q1-E1-F2

US-09-960-352-12478

Query Match 90.1%; Score 73; DB 10; Length 434;
Best Local Similarity 93.8%; Pred. No. 1.5e-17;
Matches 76; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAACGCCCAACAATGTCCTCGCCACACCCAGTGGCATGGCCGTCT 60
|||||
Db 36 GCCACATCAGACTGGCTCAATGCCCAACAATGTACCTGACCCCAAGTGGCATGGCCATCT 95
|||||

QY 61 CAAGAAGGACAGAAATCCCGAGC 81
|||||
Db 96 CAAGAAGGACAGAAATCCCGAGC 116
|||||

RESULT 9

US-09-960-352-1818

; Sequence 1818, Application US/09960352

; Patent No. US20020137139A1

; GENERAL INFORMATION:

; APPLICANT: Warren, Wesley C.

; APPLICANT: Tao, Nengbing

; APPLICANT: Byatt, John C.

; APPLICANT: Mathialagan, Nagappan

; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION

; FILE REFERENCE: 16511.006/37-21(10298)C

; CURRENT APPLICATION NUMBER: US/09/960,352

; CURRENT FILING DATE: 2001-09-24

; NUMBER OF SEQ ID NOS: 15112

; SEQ ID NO 1818

; LENGTH: 242

; TYPE: DNA

; ORGANISM: Bos taurus

; OTHER INFORMATION: Clone ID: 08-LIB34-052-Q1-E1-B11

US-09-960-352-1818

```
RESULT 2
US-09-736-457-93
; Sequence 93, Application US/09736457
; Patent No. US20020168637A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodges, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Lijun
; APPLICANT: Wang, Aijun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C15
; CURRENT APPLICATION NUMBER: US/09/736,457
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 1864
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 93
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-736-457-93

Query Match      98.0%; Score 79.4; DB 10; Length 531;
Best Local Similarity 98.8%; Pred. No. 6.3e-20;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAGCCGACACAAATGTCCTGCGCAACCCAGTGGCGCGTCT 60
    |||||
Db 157 GCCACATCAGACTGGCTCAGCCGACACAAATGTCCTGCGCAACCCAGTGGCGCGTCT 216

QY 61 CAAGAAGGACAGATCCCGCAG 81
    |||||
Db 217 CAAGAAGGACAGATCCCGCAG 237

RESULT 3
US-09-902-941-93
; Sequence 93, Application US/09902941
; Patent No. US20020172952A1
; GENERAL INFORMATION:
; APPLICANT: Henderson, Robert A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Johnson, Jeffrey C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Marnierakis, Margarita
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: McNabb, Andria
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C17
; CURRENT APPLICATION NUMBER: US/09/902,941
; CURRENT FILING DATE: 2001-07-10
; NUMBER OF SEQ ID NOS: 2002
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 93
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-902-941-93

Query Match      98.0%; Score 79.4; DB 10; Length 531;
Best Local Similarity 98.8%; Pred. No. 6.3e-20;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAGCCGACACAAATGTCCTGCGCAACCCAGTGGCGCGTCT 60
    |||||
Db 157 GCCACATCAGACTGGCTCAGCCGACACAAATGTCCTGCGCAACCCAGTGGCGCGTCT 216

QY 61 CAAGAAGGACAGATCCCGCAG 81
    |||||
Db 217 CAAGAAGGACAGATCCCGCAG 237

RESULT 4
US-09-849-626-93
; Sequence 93, Application US/09849626
; Publication No. US20020197669A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya
; APPLICANT: Fanger, Gary
; APPLICANT: Wang, Aijun
; APPLICANT: Wang, Tongtong
; APPLICANT: Switzer, Anne
; APPLICANT: McNeill, Patricia
; APPLICANT: Clapper, Jonathan
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C16
; CURRENT APPLICATION NUMBER: US/09/849,626
; CURRENT FILING DATE: 2001-05-03
; NUMBER OF SEQ ID NOS: 1926
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 93
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-849-626-93

Query Match      98.0%; Score 79.4; DB 10; Length 531;
Best Local Similarity 98.8%; Pred. No. 6.3e-20;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAGCCGACACAAATGTCCTGCGCAACCCAGTGGCGCGTCT 60
    |||||
Db 157 GCCACATCAGACTGGCTCAGCCGACACAAATGTCCTGCGCAACCCAGTGGCGCGTCT 216

QY 61 CAAGAAGGACAGATCCCGCAG 81
    |||||
Db 217 CAAGAAGGACAGATCCCGCAG 237

RESULT 5
US-09-476-300-93
; Sequence 93, Application US/09476300
; Publication No. US20030125245A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C3
; CURRENT APPLICATION NUMBER: US/09/476,300
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 785
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 93
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-476-300-93

Query Match      98.0%; Score 79.4; DB 11; Length 531;
Best Local Similarity 98.8%; Pred. No. 6.3e-20;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAGCCGACACAAATGTCCTGCGCAACCCAGTGGCGCGTCT 60
    |||||
Db 157 GCCACATCAGACTGGCTCAGCCGACACAAATGTCCTGCGCAACCCAGTGGCGCGTCT 216

QY 61 CAAGAAGGACAGATCCCGCAG 81
    |||||
Db 217 CAAGAAGGACAGATCCCGCAG 237

RESULT 4
US-09-849-626-93
; Sequence 93, Application US/09849626
; Publication No. US20020197669A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya
; APPLICANT: Fanger, Gary
; APPLICANT: Wang, Aijun
; APPLICANT: Wang, Tongtong
; APPLICANT: Switzer, Anne
; APPLICANT: McNeill, Patricia
; APPLICANT: Clapper, Jonathan
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C16
; CURRENT APPLICATION NUMBER: US/09/849,626
; CURRENT FILING DATE: 2001-05-03
; NUMBER OF SEQ ID NOS: 1926
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 93
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-849-626-93

Query Match      98.0%; Score 79.4; DB 10; Length 531;
Best Local Similarity 98.8%; Pred. No. 6.3e-20;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAGCCGACACAAATGTCCTGCGCAACCCAGTGGCGCGTCT 60
    |||||
Db 157 GCCACATCAGACTGGCTCAGCCGACACAAATGTCCTGCGCAACCCAGTGGCGCGTCT 216

QY 61 CAAGAAGGACAGATCCCGCAG 81
    |||||
Db 217 CAAGAAGGACAGATCCCGCAG 237

RESULT 5
US-09-476-300-93
; Sequence 93, Application US/09476300
; Publication No. US20030125245A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C3
; CURRENT APPLICATION NUMBER: US/09/476,300
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 785
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 93
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-476-300-93

Query Match      98.0%; Score 79.4; DB 11; Length 531;
Best Local Similarity 98.8%; Pred. No. 6.3e-20;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 2003, 00:27:37 ; Search time 151 Seconds
(without alignments)
1302.306 Million cell updates/sec

Title: US-09-585-077c-1_COPY_4300_4380
Perfect score: 81
Sequence: 1 gccacatcagactggctcaa.....aagaaggacagatccaccg 81

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1632420 seqs, 1213878141 residues

Total number of hits satisfying chosen parameters: 3264840

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	81	100.0	5215	10	US-09-880-107-1769	Sequence 1769, Ap
2	79.4	98.0	531	10	US-09-736-457-93	Sequence 93, Appl
3	79.4	98.0	531	10	US-09-902-941-93	Sequence 93, Appl
4	79.4	98.0	531	10	US-09-849-626-93	Sequence 93, Appl
5	79.4	98.0	531	11	US-09-476-300-93	Sequence 93, Appl
6	79.4	98.0	531	14	US-10-017-754-93	Sequence 93, Appl
7	79.4	98.0	5772	12	US-10-101-510-553	Sequence 553, App
8	73	90.1	434	10	US-09-960-352-12478	Sequence 12478, A
9	40.2	49.6	242	10	US-09-960-352-1818	Sequence 1818, Ap
10	30.6	37.8	31	9	US-09-801-274-1768	Sequence 1768, Ap
11	29.4	36.3	3591	11	US-09-764-891-5507	Sequence 5507, Ap
12	27.4	33.8	2193	13	US-10-027-832-97735	Sequence 97735, A
13	27.4	33.8	2193	13	US-10-027-832-97735	Sequence 97735, A
14	27.4	33.8	6471	14	US-10-207-655-103660	Sequence 103660, A
15	26.4	32.6	2168	13	US-10-068-674-1	Sequence 1, Appl
16	26	32.1	375	14	US-10-106-698-3792	Sequence 3792, Ap

c 17	26	32.1	987	10	US-09-738-973-289	Sequence 289, App
c 18	26	32.1	987	10	US-09-854-133-289	Sequence 289, App
c 19	26	32.1	987	14	US-10-144-649A-289	Sequence 289, App
c 20	25.6	31.6	2385	10	US-09-947-953-1	Sequence 1, Appl
c 21	25.4	31.4	288	10	US-09-983-965-4858	Sequence 4858, Ap
c 22	25.4	31.4	26225	9	US-09-764-869-1276	Sequence 1276, Ap
c 23	25.4	31.4	26225	14	US-10-091-504-1276	Sequence 1276, Ap
c 24	25	30.9	331	10	US-09-783-590-7357	Sequence 7357, Ap
c 25	25	30.9	414	10	US-09-833-381-1254	Sequence 1254, Ap
c 26	24.8	30.6	284	10	US-09-833-381-417	Sequence 417, App
c 27	24.8	30.6	957	13	US-10-027-632-163409	Sequence 163409, A
c 28	24.8	30.6	969	10	US-09-280-197-13	Sequence 13, Appl
c 29	24.8	30.6	3276	10	US-09-280-197-4	Sequence 4, Appl
c 30	24.8	30.6	3276	10	US-09-423-126-8	Sequence 8, Appl
c 31	24.8	30.6	4851	10	US-09-712-363-116	Sequence 116, App
c 32	24.6	30.4	723	13	US-10-027-632-165578	Sequence 165578, A
c 33	24.4	30.1	1702	10	US-09-880-107-3680	Sequence 3680, Ap
c 34	24.2	29.9	576	13	US-10-027-632-45850	Sequence 45850, A
c 35	24.2	29.9	576	13	US-10-027-632-61618	Sequence 61618, A
c 36	24.2	29.9	576	13	US-10-027-632-61619	Sequence 61619, A
c 37	24.2	29.9	576	13	US-10-027-632-107709	Sequence 107709, A
c 38	24.2	29.9	576	13	US-10-027-632-107710	Sequence 107710, A
c 39	24.2	29.9	681	13	US-10-027-632-142706	Sequence 142706, A
c 40	24.2	29.9	681	13	US-10-027-632-142707	Sequence 142707, A
c 41	24.2	29.9	1281	14	US-10-156-761-2272	Sequence 2272, Ap
c 42	24.2	29.9	9025608	14	US-10-156-761-1	Sequence 1, Appl
c 43	24	29.6	469	11	US-09-318-993-2365	Sequence 2365, Ap
c 44	24	29.6	870	13	US-10-027-632-159815	Sequence 159815, A
c 45	24	29.6	1263	10	US-09-738-626-2281	Sequence 2281, Ap

ALIGNMENTS

RESULT 1

US-09-880-107-1769
; Sequence 1769, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; PRIOR FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1769
; LENGTH: 5215
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 D90282
US-09-880-107-1769

Query Match	100.0%	Score 81;	DB 10;	Length 5215;
Best Local Similarity	100.0%	Pred. No. 2.2e+20;	Mismatches 0;	Indels 0;
Matches 81;	Conservative 0;			
Qy	1	GCCATCATGACTGGCTCAACGCCCAACATGTCCTGCCAACCCAGTGGCGCTCT	60	
Db	4295	GCCATCATGACTGGCTCAACGCCCAACATGTCCTGCCAACCCAGTGGCGCTCT	4354	
Qy	61	CAAGAAGACAGATCCAGC	81	
Db	4355	CAAGAAGACAGATCCAGC	4375	

Query Match 30.6%; Score 24.8; DB 3; Length 3276;
Best Local Similarity 57.9%; Pred.No.16;
Matches 44; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
QY 4 ACATCAGACTGGCTCAACGCCACAAATGTCCTGCCCAACCCAGTGGCATGGCCGCTCTCAA 63
Db 1362 ACCITCGACAAATCTCTTGACACAGAGATGTTATACCAACCCTCTGGTCTTGTCTCTCAA 1303
QY 64 GAAGGACAGAAATCCCA 79
Db 1302 CAAGGACAAATCCCA 1287

Search completed: September 13, 2003, 00:29:22
Job time : 64.5 secs

[illegible]

```
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Lijun
; APPLICANT: Wang, Aijun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C15
; CURRENT APPLICATION NUMBER: US/09/736,457
; CURRENT FILING DATE: 2000-12-13
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 93
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-736-457-93

Query Match      98.0%; Score 79.4; DB 4; Length 531;
Best Local Similarity 98.8%; Pred. No. 1.3e-18;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACATCAGATGGCTCAAGCCCAACAATGTCCTGGCCAAACCCAGTGGCATGGCGTCT 60
    |||||
Db 157 GCCACATCAGATGGCTCAAGCCCAACAATGTCCTGGCCAAACCCAGTGGCATGGCGTCT 216

QY 61 CAAGAGGACAGAAATCCAGC 81
    |||||
Db 217 CAAGAGGACAGAAATCCAGC 237

RESULT 7
US-09-323-472A-3
; Sequence 3, Application US/09323472A
; Patent No. 6346382
; GENERAL INFORMATION:
; APPLICANT: Summar, Marshall
; APPLICANT: Christman, Brian
; TITLE OF INVENTION: HUMAN CARBAMYL PHOSPHATE SYNTHETASE I POLYMORPHISM AND DIAGNOSIS
; TITLE OF INVENTION: THERAPEUTIC METHODS RELATING THERETO
; FILE REFERENCE: 1242/19
; CURRENT APPLICATION NUMBER: US/09/323,472A
; CURRENT FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 5761
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (124)..(4626)
US-09-323-472A-3

Query Match      98.0%; Score 79.4; DB 4; Length 5761;
Best Local Similarity 98.8%; Pred. No. 2.4e-18;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACATCAGATGGCTCAAGCCCAACAATGTCCTGGCCAAACCCAGTGGCATGGCGTCT 60
    |||||
Db 4300 GCCACATCAGATGGCTCAAGCCCAACAATGTCCTGGCCAAACCCAGTGGCATGGCGTCT 4359

QY 61 CAAGAGGACAGAAATCCAGC 81
    |||||
Db 4360 CAAGAGGACAGAAATCCAGC 4380

RESULT 8
US-08-633-879C-1/c
; Sequence 1, Application US/08633879C
; Patent No. 5928922
; GENERAL INFORMATION:
; APPLICANT: Kivirikko, Kari I.
; APPLICANT: Pihlajaniemi, Taina
```

```
; APPLICANT: Helaakoski, Tarja I.
; APPLICANT: Annunen, Pia P.
; APPLICANT: Nissi, Ritva K.
; APPLICANT: No. 5928922elainen, Minna K.
; TITLE OF INVENTION: '2 SUBUNIT OF PROLYL-4-HYDROXYLASE
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING SUCH SUBUNIT AND
; TITLE OF INVENTION: METHODS FOR PRODUCING THE SAME
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds, LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/633,879C
; FILING DATE: 10-APR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Abrams, Samuel B
; REGISTRATION NUMBER: 30,605
; REFERENCE/DOCKET NUMBER: 8389-0041-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2168 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 151..1761
; OTHER INFORMATION:
US-08-633-879C-1

Query Match      32.6%; Score 26.4; DB 2; Length 2168;
Best Local Similarity 61.8%; Pred. No. 4.1;
Matches 42; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 9 AGACTGGCTCAAGCCCAACAATGTCCTGGCCAAACCCAGTGGCATGGCGTCTCAAGAAG 68
    |||||
Db 2005 AGTCTGGCTTACCTACCGCAGCCATGCCAAGACACCCCGAGCCAGGCGAGGTGTAAGCTGG 1946

QY 69 ACAGAATC 76
    |||||
Db 1945 TCAGAGTC 1938

RESULT 9
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
```


; SEQ ID NO 11
; LENGTH: 5761
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (124)..(4626)
US-09-323-472A-11

Query Match 100.0%; Score 81; DB 4; Length 5761;
Best Local Similarity 100.0%; Pred. No. 6.6e-19; Mismatches 0; Indels 0; Gaps 0;
Matches 81; Conservative 0;

QY 1 GCCATCAGACTGGCTCAACGCCAACCAATGTCCTGCCAACCCAGTGGCATGGCCGTCT 60
|||||
Db 4300 GCCATCAGACTGGCTCAACGCCAACCAATGTCCTGCCAACCCAGTGGCATGGCCGTCT 4359

QY 61 CAAGAAGGACAGATCCCGC 81
|||||
Db 4360 CAAGAAGGACAGATCCCGC 4380

RESULT 3

US-09-323-472A-13
; Sequence 13, Application US/09323472A
; Patent No. 6346382
; GENERAL INFORMATION:

; APPLICANT: Summar, Marshall
; APPLICANT: Christman, Brian
; TITLE OF INVENTION: HUMAN CARBAMYL PHOSPHATE SYNTHETASE I POLYMORPHISM AND DIAGNOSTIC
; TITLE OF INVENTION: THERAPEUTIC METHODS RELATING THERETO
; FILE REFERENCE: 1242/19
; CURRENT APPLICATION NUMBER: US/09/323,472A
; CURRENT FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
; LENGTH: 5762
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (124)..(4626)
US-09-323-472A-13

Query Match 100.0%; Score 81; DB 4; Length 5762;
Best Local Similarity 100.0%; Pred. No. 6.6e-19; Mismatches 0; Indels 0; Gaps 0;
Matches 81; Conservative 0;

QY 1 GCCATCAGACTGGCTCAACGCCAACCAATGTCCTGCCAACCCAGTGGCATGGCCGTCT 60
|||||
Db 4300 GCCATCAGACTGGCTCAACGCCAACCAATGTCCTGCCAACCCAGTGGCATGGCCGTCT 4359

QY 61 CAAGAAGGACAGATCCCGC 81
|||||
Db 4360 CAAGAAGGACAGATCCCGC 4380

RESULT 4

US-09-323-472A-5
; Sequence 5, Application US/09323472A
; Patent No. 6346382
; GENERAL INFORMATION:

; APPLICANT: Summar, Marshall
; APPLICANT: Christman, Brian
; TITLE OF INVENTION: HUMAN CARBAMYL PHOSPHATE SYNTHETASE I POLYMORPHISM AND DIAGNOSTIC
; TITLE OF INVENTION: THERAPEUTIC METHODS RELATING THERETO
; FILE REFERENCE: 1242/19
; CURRENT APPLICATION NUMBER: US/09/323,472A
; CURRENT FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5

; LENGTH: 495
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: exon
; LOCATION: (111)..(224)
; OTHER INFORMATION: n is G or A or C or T/U
US-09-323-472A-5

Query Match 98.0%; Score 79.4; DB 4; Length 495;
Best Local Similarity 98.8%; Pred. No. 1.3e-18; Mismatches 1; Indels 0; Gaps 0;
Matches 80; Conservative 0;

QY 1 GCCATCAGACTGGCTCAACGCCAACCAATGTCCTGCCAACCCAGTGGCATGGCCGTCT 60
|||||
Db 126 GCCATCAGACTGGCTCAACGCCAACCAATGTCCTGCCAACCCAGTGGCATGGCCGTCT 185

QY 61 CAAGAAGGACAGATCCCGC 81
|||||
Db 186 CAAGAAGGACAGATCCCGC 206

RESULT 5

US-09-702-705-93
; Sequence 93, Application US/09702705
; Patent No. 6504010
; GENERAL INFORMATION:

; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C14
; CURRENT APPLICATION NUMBER: US/09/702,705
; CURRENT FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 1833
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 93
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-702-705-93

Query Match 98.0%; Score 79.4; DB 4; Length 531;
Best Local Similarity 98.8%; Pred. No. 1.3e-18; Mismatches 1; Indels 0; Gaps 0;
Matches 80; Conservative 0;

QY 1 GCCATCAGACTGGCTCAACGCCAACCAATGTCCTGCCAACCCAGTGGCATGGCCGTCT 60
|||||
Db 157 GCCATCAGACTGGCTCAACGCCAACCAATGTCCTGCCAACCCAGTGGCATGGCCGTCT 216

QY 61 CAAGAAGGACAGATCCCGC 81
|||||
Db 217 CAAGAAGGACAGATCCCGC 237

RESULT 6

US-09-736-457-93
; Sequence 93, Application US/09736457
; Patent No. 6509448
; GENERAL INFORMATION:

; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2003, 22:27:31 ; Search time 44.5 Seconds
(without alignments)
803.416 Million cell updates/sec

Title: US-09-585-077C-1_COPY_4300_4380
Perfect score: 81
Sequence: 1 gccacatcagctggtctcaacgagacagaatccaccgc 81

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, NA: *
1: /cgn2_6/prodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/prodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/prodata/2/ina/5A_COMB.seq: *
4: /cgn2_6/prodata/2/ina/5B_COMB.seq: *
5: /cgn2_6/prodata/2/ina/PCTUS_COMB.seq: *
6: /cgn2_6/prodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	81	100.0	5761	4	US-09-323-472A-1
2	81	100.0	5761	4	US-09-323-472A-11
3	81	100.0	5762	4	US-09-323-472A-13
4	79.4	98.0	495	4	US-09-323-472A-5
5	79.4	98.0	531	4	US-09-702-705-93
6	79.4	98.0	531	4	US-09-736-457-93
7	79.4	98.0	5761	4	US-09-323-472A-3
8	26.4	32.6	2168	2	US-08-633-879C-1
9	26.2	32.3	4403765	3	US-09-103-840A-2
10	26.2	32.3	4411529	3	US-09-103-840A-1
11	26	32.1	987	4	US-09-370-838-289
12	25.6	31.6	458	3	US-09-067-782A-6
13	25.6	31.6	1056	4	US-09-016-434-1308
14	24.8	30.6	654	4	US-09-252-991A-12606
15	24.8	30.6	3276	3	US-08-633-768A-4
16	24	29.6	432	4	US-09-252-991A-16133
17	24	29.6	438	4	US-09-252-991A-8292
18	23.8	29.4	1703	3	US-09-135-021-77
19	23.8	29.4	1703	3	US-09-135-020-3
20	23.8	29.4	1703	3	US-09-135-010A-3
21	23.8	29.4	1703	4	US-09-444-871-3
22	23.8	29.4	1703	4	US-09-597-735-3
23	23.8	29.4	1703	4	US-09-444-295-3
24	23.8	29.4	1703	4	US-09-597-732-3
25	23.8	29.4	1703	4	US-09-597-731-3
26	23.6	29.1	1056	3	US-09-067-782A-1
27	23.6	29.1	1520	5	PCT-US93-07213-8

c	28	23.6	29.1	1820	2	US-08-828-922-2	Sequence 2, Appli
c	29	23.6	29.1	2219	5	PCT-US93-07213-4	Sequence 4, Appli
c	30	23.6	29.1	2870	5	PCT-US93-07213-3	Sequence 3, Appli
c	31	23.6	29.1	3102	5	PCT-US93-07213-1	Sequence 1, Appli
c	32	23.6	29.1	152331	3	US-09-128-155-16	Sequence 16, Appli
c	33	23.4	28.9	519	4	US-09-844-634-19	Sequence 19, Appli
c	34	23.4	28.9	1017	4	US-09-252-991A-11573	Sequence 11573, A
c	35	23.4	28.9	1056	4	US-09-252-991A-11806	Sequence 11806, A
c	36	23.4	28.9	1140	4	US-09-252-991A-11515	Sequence 11515, A
c	37	23.4	28.9	3683	4	US-09-844-634-3	Sequence 3, Appli
c	38	23.4	28.9	15602	4	US-09-844-634-17	Sequence 17, Appli
c	39	23.2	28.6	1900	4	US-09-608-285A-47	Sequence 47, Appli
c	40	23.2	28.6	1900	4	US-09-557-800C-49	Sequence 49, Appli
c	41	23.2	28.6	2294	4	US-09-608-285A-49	Sequence 49, Appli
c	42	23.2	28.6	2294	4	US-09-557-800C-49	Sequence 49, Appli
c	43	23.2	28.6	2371	4	US-09-608-285A-46	Sequence 46, Appli
c	44	23.2	28.6	2371	4	US-09-557-800C-46	Sequence 46, Appli
c	45	23.2	28.6	2497	4	US-09-608-285A-51	Sequence 51, Appli

ALIGNMENTS

RESULT 1
US-09-323-472A-1
; Sequence 1, Application US/09323472A
; Patent No. 6346382
; GENERAL INFORMATION:
; APPLICANT: Summar, Marshall
; TITLE OF INVENTION: HUMAN CARBAMYL PHOSPHATE SYNTHETASE I POLYMORPHISM AND DIAGN
; TITLE OF INVENTION: THERAPEUTIC METHODS RELATING THERETO
; FILE REFERENCE: 1242/19
; CURRENT APPLICATION NUMBER: US/09/323,472A
; CURRENT FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 5761
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (124)..(4626)
US-09-323-472A-1

Query Match 100.0%; Score 81; DB 4; Length 5761;
Best Local Similarity 100.0%; Pred. No. 6.6e-19;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCCACATCAGACTGGCTCAACGCCCAACAATGCTCGTCCCAACCCAGTGGCGCTCT 60
Db 4300 GCCACATCAGACTGGCTCAACGCCCAACAATGCTCGTCCCAACCCAGTGGCGCTCT 4359
QY 61 CAAGAAGGACAGATCCCAAGC 81
Db 4360 CAAGAAGGACAGATCCCAAGC 4380

RESULT 2
US-09-323-472A-11
; Sequence 11, Application US/09323472A
; Patent No. 6346382
; GENERAL INFORMATION:
; APPLICANT: Summar, Marshall
; APPLICANT: Christman, Brian
; TITLE OF INVENTION: HUMAN CARBAMYL PHOSPHATE SYNTHETASE I POLYMORPHISM AND DIAGN
; TITLE OF INVENTION: THERAPEUTIC METHODS RELATING THERETO
; FILE REFERENCE: 1242/19
; CURRENT APPLICATION NUMBER: US/09/323,472A
; CURRENT FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0

```

mRNA sequence.
ACCESSION      BG533520
VERSION        BG533520.1  GI:13525060
KEYWORDS       EST.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 949)
AUTHORS        NIH-MGC http://mgc.nci.nih.gov/.
TITLE          National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL        Unpublished
COMMENT        Contact: Robert Strausberg, Ph.D.
               Email: cgabbs-r@mail.nih.gov
               Tissue Procurement: CLONETECH Laboratories, Inc.
               cDNA Library Preparation: CLONETECH Laboratories, Inc.
               cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
               DNA Sequencing by: Incyte Genomics, Inc.
               Clone distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               http://image.llnl.gov
               Plate: LCM918 row: f column: 09
               High quality sequence stop: 740.
               Location/Qualifiers
                 1..949
                   /organism="Homo sapiens"
                   /mol_type="mRNA"
                   /db_xref="taxon:9606"
                   /clone="IMAGE:4071944"
                   /lab_host="DH10B (T1 phage-resistant)"
                   /clone_lib="NIH_MGC_76"
                   /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
                   SfiI (ggcccattcgcc); Site_2: SfiI (ggccattatgcc); 5' and
                   3' adaptors were used in cloning as follows: 5' adaptor
                   sequence: 5'-CAGGCCATTATGCC-3' and 3' adaptor sequence:
                   5'-ATTCTAGAGCGCGGCGGACATG-dt(30)BN-3' (where B = A,
                   C, or G and N = A, C, G, or T). Average insert size 1.85
                   kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
                   by PCR. This library was enriched for full-length clones
                   and was constructed by Clontech Laboratories (Palo Alto,
                   CA). Note: this is a NIH_MGC Library."
BASE COUNT     271 a 224 c 181 g 272 t 1 others
ORIGIN

Query Match      98.0%; Score 79.4; DB 10; Length 949;
Best Local Similarity 98.8%; Pred. No. 8.6e-15;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTCGCCACCCAGTGGCATGGCGGTCT 60
    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTCGCCACCCAGTGGCATGGCGGTCT 240

QY 61 CAGAGAGGACAGATCCCGAGC 81
    |||||||||||||||||||||||
Db 241 CAGAGAGGACAGATCCCGAGC 261

RESULT 15
AV656423
LOCUS          AV656423 362 bp mRNA linear EST 16-JAN-2002
DEFINITION    AV656423 GLC Homo sapiens cDNA clone GLCERB10 3', mRNA sequence.
ACCESSION     AV656423
VERSION       AV656423.1 GI:9877437
KEYWORDS      EST.
SOURCE        Homo sapiens (human)
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 362)
AUTHORS        Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
               Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
               Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
               Hu,G., Gu,J., Chen,Z. and Han,Z.

```

```

TITLE          Insight into hepatocellular carcinogenesis at transcriptome level
               by comparing gene expression profiles of hepatocellular carcinoma
               with those of corresponding noncancerous liver
               Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
JOURNAL        21625106
MEDLINE        11752456
PUBMED         11752456
COMMENT        Contact: Zeguang Han
               Chinese National Human Genome Center at Shanghai
               351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
               201203, P. R. China
               Tel: 86-21-50801919(ex.45)
               Fax: 86-21-50801922
               Email: hanzg@chgc.sh.cn
               This clone is available at CHGC in Shanghai.
FEATURES       Location/Qualifiers
                 1..362
                   /organism="Homo sapiens"
                   /mol_type="mRNA"
                   /db_xref="taxon:9606"
                   /clone="GLCERB10"
                   /tissue_type="corresponding non cancerous liver tissue"
                   /dev_stage="Adult"
                   /lab_host="SOLR"
                   /clone_lib="GLC"
                   /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                   XhoI"
BASE COUNT     106 a 92 c 71 g 93 t
ORIGIN

Query Match      96.8%; Score 78.4; DB 9; Length 362;
Best Local Similarity 98.8%; Pred. No. 1.2e-14;
Matches 79; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCACATCAGACTGGCTCAACGCCAACAAATGTCCTCGCCACCCAGTGGCATGGCGGTCTC 61
    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 CCACATCAGACTGGCTCAACGCCAACAAATGTCCTCGCCACCCAGTGGCATGGCGGTCTC 60

QY 62 AAGAAGGACAGATCCCGAGC 81
    |||||||||||||||||||||||
Db 61 AAGAAGGACAGATCCCGAGC 80

Search completed: September 13, 2003, 00:27:26
Job time : 1968 secs

```

```

FEATURES
  source
    Location/Qualifiers
      1. .637
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="GLC"
        /tissue_type="corresponding non cancerous liver tissue"
        /dev_stage="Adult"
        /lab_host="SOLR"
        /clone_lib="GLC"
        /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"
BASE COUNT      179 a 148 c 122 g 185 t
ORIGIN
  Query Match      98.0%; Score 79.4; DB 9; Length 636;
  Best Local Similarity 98.8%; Pred. No. 7.4e-15;
  Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAAGCCCAACAATGTCCTGCCAACCCAGTGGCATGCCGTCT 60
    |||||||
Db 5 GCCACATCAGACTGGCTCAAGCCCAACAATGTCCTGCCAACCCAGTGGCATGCCGTCT 64
    |||||||
QY 61 CAAGAAGACAGATCCAGC 81
    |||||||
Db 65 CAAGAAGACAGATCCAGC 85
    |||||||

RESULT 13
AL046242/c
LOCUS
DEFINITION DKFZp434E087.sl 434 (synonym: htes3) Homo sapiens cDNA clone
ACCESSION AL046242
VERSION AL046242.1 GI:5434326
SOURCE EST.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 646)
AUTHORS Koehrer,K., Beyer,A., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
TITLE EST (Koehrer, et al.)
JOURNAL Unpublished
COMMENT Contact: Koehrer K
MIPS
Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 3' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by BMF (Biomedical Research Center at the Charite,
Berlin/Germany) within the cDNA sequencing consortium of the German
Genome Project
x1 sequence also available.
This clone (DKFZp434E087) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES
  source
    1. .646
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="DKFZp434E087"
      /tissue_type="testis"
      /dev_stage="adult"
      /lab_host="DH10B"
      /clone_lib="434 (synonym: htes3)"
      /note="Vector: pSport1; Site_1: NotI; Site_2: SalI"
BASE COUNT      171 a 128 c 156 g 190 t
ORIGIN
  Query Match      98.0%; Score 79.4; DB 9; Length 646;
  Best Local Similarity 98.8%; Pred. No. 7.4e-15;
  Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAAGCCCAACAATGTCCTGCCAACCCAGTGGCATGCCGTCT 60
    |||||||
Db 469 GCCACATCAGACTGGCTCAAGCCCAACAATGTCCTGCCAACCCAGTGGCATGCCGTCT 410
    |||||||
QY 61 CAAGAAGACAGATCCAGC 81
    |||||||
Db 409 CAAGAAGACAGATCCAGC 389
    |||||||

RESULT 14
BG533520
LOCUS
DEFINITION NIH_MGC_76 Homo sapiens cDNA clone IMAGE:407194 5',

```

```

FEATURES
  source
    Location/Qualifiers
      1. .637
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="GLC"
        /tissue_type="corresponding non cancerous liver tissue"
        /dev_stage="Adult"
        /lab_host="SOLR"
        /clone_lib="GLC"
        /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"
BASE COUNT      179 a 148 c 122 g 185 t
ORIGIN
  Query Match      98.0%; Score 79.4; DB 9; Length 636;
  Best Local Similarity 98.8%; Pred. No. 7.4e-15;
  Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAAGCCCAACAATGTCCTGCCAACCCAGTGGCATGCCGTCT 60
    |||||||
Db 59 GCCACATCAGACTGGCTCAAGCCCAACAATGTCCTGCCAACCCAGTGGCATGCCGTCT 118
    |||||||
QY 61 CAAGAAGACAGATCCAGC 81
    |||||||
Db 119 CAAGAAGACAGATCCAGC 139
    |||||||

RESULT 12
AV683932
LOCUS
DEFINITION AV683932 GK Homo sapiens cDNA clone GKCID08 5', mRNA sequence.
ACCESSION AV683932
VERSION AV683932.1 GI:10285795
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 637)
AUTHORS Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
Hu,G., Gu,J., Chen,Z. and Han,Z.
TITLE Insight into hepatocellular carcinogenesis at transcriptome level
by comparing gene expression profiles of hepatocellular carcinoma
with those of corresponding noncancerous liver
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
MEDLINE 21625106
PUBMED 11752456
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzq@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
  source
    1. .637
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="GKCID08"
      /tissue_type="hepatocellular carcinoma"
      /dev_stage="Adult"
      /lab_host="SOLR"
      /clone_lib="GKC"
      /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"
BASE COUNT      173 a 147 c 118 g 196 t
ORIGIN

```

RESULT 9

AV661502 564 bp mRNA linear EST 16-JAN-2002
 LOCUS AV661502 GLC Homo sapiens cDNA clone GLCGB04 3', mRNA sequence.
 DEFINITION AV661502
 ACCESSION AV661502
 VERSION AV661502.1 GI:9882516
 KEYWORDS EST.

SOURCE

Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

1 (bases 1 to 564)
 Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
 Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
 Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
 Hu,G., Gu,J., Chen,Z., and Han,Z.

Insight into hepatocellular carcinogenesis at transcriptome level
 by comparing gene expression profiles of hepatocellular carcinoma
 with those of corresponding noncancerous liver

Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)

21625106

11752456

Contact: Zequang Han

Chinese National Human Genome Center at Shanghai

351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai

201203, P. R. China

Tel: 86-21-50801919(ex.45)

Fax: 86-21-50801922

Email: hanzg@chgc.sh.cn

This clone is available at CHGC in Shanghai.

FEATURES

source

Location/Qualifiers
 1..564
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="GLCGB04"
 /tissue_type="corresponding non cancerous liver tissue"
 /dev_stage="Adult"
 /lab_host="SOLR"
 /clone_lib="GLC"
 /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
 XhoI"

BASE COUNT 140 a 140 c 129 g 155 t

ORIGIN

Query Match 98.0%; Score 79.4; DB 9; Length 564;
 Best Local Similarity 98.8%; Pred. No. 7e-15;
 Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTGCCAACCCAGTGGCGCTCT 60

DB 453 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTGCCAACCCAGTGGCGCTCT 512

QY 61 CAAGAAGGACAGAAATCCAGC 81

DB 513 CAAGAAGGACAGAAATCCAGC 533

RESULT 10

CB153827 625 bp mRNA linear EST 29-JAN-2003
 LOCUS CB153827 GLC Homo sapiens cDNA clone L17N670205-7-B10
 DEFINITION CB153827
 ACCESSION CB153827
 VERSION CB153827.1 GI:28138823
 KEYWORDS EST.

SOURCE

Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

1 (bases 1 to 625)
 Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
 Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and

Kim,Y.S.

21C Frontier Korean EST Project 2001

Unpublished

Contact: Kim YS

Genome Research Center

Korea Research Institute of Bioscience & Biotechnology

52 Eoeun-dong yuseong-gu, Daejeon 305-333, South Korea

Tel: +82-42-860-4470

Fax: +82-42-860-4409

Email: yongsung@mail.kribb.re.kr

Plate: 7 row: B column: 10

High quality sequence stop: 625.

FEATURES

source

Location/Qualifiers
 1..625
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="L17N670205-7-B10"
 /sex="F"
 /lab_host="Top10F"
 /clone_lib="L17N670205"
 /note="Organ: Liver; Vector: pT73-Pac; Site_1: EcoRI;
 Site_2: NotI: The library was contributed by the Soares
 laboratory and it was constructed as described by Bonaldo,
 X.F., Lennon, G. and Soares, M.B. (1996), Genome Research
 6(5): 791-806. RNA was prepared from harvested cell
 culture."

BASE COUNT 178 a 159 c 134 g 154 t

ORIGIN

Query Match 98.0%; Score 79.4; DB 14; Length 625;
 Best Local Similarity 98.8%; Pred. No. 7.3e-15;
 Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT 11

AV654671 636 bp mRNA linear EST 15-JAN-2002
 LOCUS AV654671 GLC Homo sapiens cDNA clone GLCDBC09 3', mRNA sequence.
 DEFINITION AV654671
 ACCESSION AV654671
 VERSION AV654671.1 GI:9875685

KEYWORDS

EST.

SOURCE

Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 636)
 Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
 Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
 Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
 Hu,G., Gu,J., Chen,Z., and Han,Z.
 Insight into hepatocellular carcinogenesis at transcriptome level
 by comparing gene expression profiles of hepatocellular carcinoma
 with those of corresponding noncancerous liver

Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)

21625106

11752456

COMMENT

Contact: Zequang Han
 Chinese National Human Genome Center at Shanghai
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
 201203, P. R. China
 Tel: 86-21-50801919(ex.45)
 Fax: 86-21-50801922
 Email: hanzg@chgc.sh.cn

This clone is available at CHGC in Shanghai.

```

Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
  source
    1..374
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="GLOCNF11"
      /tissue_type="corresponding non cancerous liver tissue"
      /dev_stage="Adult"
      /lab_host="SOLP"
      /clone_lib="GLC"
      /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
      XhoI"
BASE COUNT      91 a 103 c 87 g 93 t
ORIGIN
Query Match      98.0%; Score 79.4; DB 9; Length 374;
Best Local Similarity 98.8%; Pred. No. 6e-15; 1; Indels 0; Gaps 0;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCATCATGAGTGGCTCAACGCCCAACAATGTCCTGCCACCCAGTGGCGGCTCT 60
    |||||||
Db 277 GCCATCATGAGTGGCTCAACGCCCAACAATGTCCTGCCACCCAGTGGCGGCTCT 336

QY 61 CAAGAGGACAGATCCCGC 81
    |||||||
Db 337 CAAGAGGACAGATCCCGC 357

RESULT 7
LOCUS      A1174790      501 bp      mRNA      linear      EST 11-NOV-1999
DEFINITION      HAZ511 Human fetal liver cDNA library Homo sapiens cDNA, mRNA
sequence.
ACCESSION      A1174790
VERSION        A1174790.1 GI:63611174
KEYWORDS      EST.
SOURCE        Homo sapiens (human)
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 501)
AUTHORS        Yu, Y., Zhang, C., Luo, L., Ouyang, S., Li, W., Wu, J., Zhou, S., Liu, M.
and He, F.
TITLE          Expression profile analysis of a human fetal liver cDNA library
JOURNAL        Unpublished
COMMENT        Contact: Chenggang Zhang
                Beijing Institute of Radiation Medicine
                27 Taiping Road, Beijing 100850, P.R.China
                Email: zhang.chenggang@hotmail.com.

FEATURES
  source
    1..501
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /tissue_type="liver"
      /dev_stage="fetal"
      /lab_host="MCL061/P3"
      /clone_lib="Human fetal liver cDNA library"
      /note="Vector: pCDNA1"
BASE COUNT      143 a 124 c 110 g 124 t
ORIGIN
Query Match      98.0%; Score 79.4; DB 9; Length 501;
Best Local Similarity 98.8%; Pred. No. 6.7e-15;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCATCATGAGTGGCTCAACGCCCAACAATGTCCTGCCACCCAGTGGCGGCTCT 60
    |||||||
Db 181 GCCATCATGAGTGGCTCAACGCCCAACAATGTCCTGCCACCCAGTGGCGGCTCT 240

```

```

QY 51 CAAGAGGACAGATCCCGC 81
    |||||||
Db 241 CAAGAGGACAGATCCCGC 261

RESULT 8
LOCUS      BM821786      540 bp      mRNA      linear      EST 06-MAR-2002
DEFINITION      K-EST0091028 S20T665307 Homo sapiens cDNA clone S20T665307-14-F07
5', mRNA sequence.
ACCESSION      BM821786
VERSION        BM821786.1 GI:19178199
KEYWORDS      EST.
SOURCE        Homo sapiens (human)
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 540)
AUTHORS        Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,
Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and
Kim, Y.S.
TITLE          21C Frontier Korean EST Project 2001
JOURNAL        Unpublished
COMMENT        Contact: Kim YS
                Genome Research Center
                Korea Research Institute of Bioscience & Biotechnology
                52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
                Tel: +82-42-860-4470
                Fax: +82-42-860-4409
                Email: yongsung@mail.kribb.re.kr
                Plate: 14 row: F column: 07
                High quality sequence stop: 540.

FEATURES
  source
    1..540
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="S20T665307-14-F07"
      /sex="M"
      /lab_host="Top10F"
      /clone_lib="S20T665307"
      /note="Organ: Stomach; Vector: pCNS; Site_1: EcoRI;
      Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
      bacterial alkaline phosphatase (BAP) and then decapped
      with tobacco acid pyrophosphatase (TAP). The decapped
      intact mRNA was ligated with DNA-RNA linker including EcoR
      I site by treatment of T4 RNA ligase and the first strand
      cDNA was synthesized from oligo dt-selected mRNA by
      priming with dt-tailed vector. The dt-tailed vector was
      adjusted to have about 60nt. The cDNA vector was
      circularized with E. coli DNA ligase after digestion of
      EcoRI which site is also included in vector. An RNA strand
      converted to a DNA strand by Okayama-Berg method. The
      obtained cDNA vectors were used for transformation of
      competent cells E. coli Top10F by electroporation method.
      The cDNA libraries constructed by this method are
      full-length enriched cDNA library."
BASE COUNT      147 a 133 c 121 g 139 t
ORIGIN
Query Match      98.0%; Score 79.4; DB 12; Length 540;
Best Local Similarity 98.8%; Pred. No. 6.9e-15;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCATCATGAGTGGCTCAACGCCCAACAATGTCCTGCCACCCAGTGGCGGCTCT 60
    |||||||
Db 329 GCCATCATGAGTGGCTCAACGCCCAACAATGTCCTGCCACCCAGTGGCGGCTCT 388

QY 61 CAAGAGGACAGATCCCGC 81
    |||||||
Db 389 CAAGAGGACAGATCCCGC 409

```

AUTHORS Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X., Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W., Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X., Hu,G., Gu,J., Chen,Z. and Han,Z.

TITLE Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)

MEDLINE 21625106

PUBMED 11752456

COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES source
1..366
location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="GKBCUC07"
/tissue_type="hepatocellular carcinoma"
/dev_stage="Adult"
/lab_host="SOLR"
/clone_lib="GKC"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 92 a 98 c 82 g 93 t 1 others

ORIGIN

Query Match 98.0%; Score 79.4; DB 9; Length 366;
Best Local Similarity 98.8%; Pred. No. 6e-15; Indels 0; Gaps 0;
Matches 80; Conservative 0; Mismatches 1;

QY 1 GCACATCAGATGGCTCAAGCCAAACAATGTCCTGCACACCACTGGCGCGTCT 60

Db 193 GCCACATCAGATGGCTCAGCGCCACATATGTCCTGCACCCAGTGCATGGCGCGTCT 252

QY 61 CAAGAAGGACAGATCCGAGC 81

Db 253 CAAGAAGGACAGATCCGAGC 273

RESULT 6

AV661012 374 bp mRNA linear EST 16-JAN-2002

LOCUS AV661012 GLC Homo sapiens cDNA clone GLOCNFI1 3', mRNA sequence.

DEFINITION AV661012

ACCESSION AV661012

VERSION AV661012.1 GI:9882026

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 374)
Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X., Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W., Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X., Hu,G., Gu,J., Chen,Z. and Han,Z.

REFERENCE Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver

AUTHORS Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)

JOURNAL 21625106

MEDLINE 11752456

PUBMED 11752456

COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China
Tel: 86-21-50801919(ex.45)

2

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2003, 22:26:16 ; Search time 1960 Seconds
(without alignments)
1004.420 Million cell updates/sec

Title: US-09-585-077c-1_COPY_4300_4380

Perfect score: 81

Sequence: 1 gccacatcagctgctcaa.....aagaaggacagatccacgc 81

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*

1: em_estba.*

2: em_esthum.*

3: em_estlin.*

4: em_estmu.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_hic.*

9: gb_est1.*

10: gb_est2.*

11: gb_hic.*

12: gb_est3.*

13: gb_est4.*

14: gb_est5.*

15: em_estfun.*

16: em_estom.*

17: em_gss_hum.*

18: em_gss_inv.*

19: em_gss_pln.*

20: em_gss_vrt.*

21: em_gss_fun.*

22: em_gss_jman.*

23: em_gss_jmus.*

24: em_gss_pro.*

25: em_gss_rod.*

26: em_gss_phg.*

27: em_gss_vri.*

28: gb_gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	81	100.0	699	12	BG761337
2	81	100.0	706	10	BG616938
3	81	100.0	843	10	BE971350
4	80	98.8	490	9	AI065054

5	79.4	98.0	366	9	AV696724
6	79.4	98.0	374	9	AV661012
7	79.4	98.0	501	9	AI174790
8	79.4	98.0	540	12	BM821786
9	79.4	98.0	564	9	AV661502
10	79.4	98.0	625	14	CB153827
11	79.4	98.0	636	9	AV654671
12	79.4	98.0	637	9	AV683932
13	79.4	98.0	646	9	AL046242
14	79.4	98.0	949	10	BG533520
15	78.4	96.8	362	9	AV656423
16	77.8	96.0	355	9	AV660978
17	77.8	96.0	626	9	AV653332
18	77.8	96.0	830	10	BG569284
19	74.6	92.1	444	9	AA883205
20	73.2	90.4	405	14	T59186
21	73	90.1	350	14	CB781273
22	73	90.1	554	12	BM432173
23	73	90.1	688	14	CB462893
24	70.2	86.7	512	12	BQ011349
25	69.8	86.2	473	9	AI786068
26	69.8	86.2	492	9	AA250015
27	69.8	86.2	524	9	AA268939
28	69.8	86.2	629	9	AA986202
29	69.8	86.2	706	14	CB950475
30	69.8	86.2	740	14	CB948817
31	69.8	86.2	791	12	BI217440
32	69.8	86.2	920	10	BF533080
33	69.8	86.2	1057	14	W29382
34	69.8	86.2	3161	11	AK028683
35	69	85.2	307	14	T77346
36	69	85.2	521	9	AA190971
37	68.8	84.9	723	12	BI247566
38	68.2	84.2	511	9	AA511517
39	67.4	83.2	541	9	AA393420
40	67.4	83.2	553	9	AA398737
41	65.4	80.7	391	9	AV655328
42	64	79.0	622	9	AI047399
43	60.4	74.6	564	14	CB120338
44	60	74.1	372	10	BF547163
45	58.4	72.1	745	14	CB594152

ALIGNMENTS

RESULT 1
BG761337
LOCUS BG761337 699 bp mRNA linear EST 15-MAY-2001
DEFINITION 602718213F1 NIH_MGC_49 Homo sapiens CDNA clone IMAGE:4858442 5', mRNA sequence.
ACCESSION BG761337
VERSION BG761337.1 GI:14071990
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 699)
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC/DCTD/DTF
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: L10M1712 row: i column: 03
High quality sequence stop: 695.

QY 61 CARGAGGACAGATCCGAC 81
Db ||||||||||||||||||
96 CARGAGGACAGATCCGAC 116

Search completed: September 12, 2003, 22:34:25
Job time : 211.5 secs

XX Toxicologically relevant human nucleotide sequence #1900.
DE
XX
XX Toxicologically relevant gene; toxicological response; gene; ss.
KW
XX
XX Homo sapiens.
OS
XX
XX WC2003016500-A2.
PN
XX
XX 27-FEB-2003.
PD
XX
XX 16-AUG-2002; 2002WO-US26514.
PF
XX
XX 16-AUG-2001; 2001US-313080P.
PR
XX
XX (PHAS-) PHASE-1 MOLECULAR TOXICOLOGY INC.
PA
XX
XX Reft RE, Dunn RT, Adkins K, Pickett GG, Kier LD, Schmeiser K;
PI
PI Allen P;
XX
XX WPI; 2003-268322/26.
DR
XX
XX Determining a toxicological response to an agent, useful for screening
PT
PT of drugs, comprises comparing the expression profile of one or more
PT human toxic response genes to a reference gene expression profile
PT indicative of toxicity -
XX
XX Claim 1; Page 447; 455pp; English.
FS
XX
XX The present invention describes a method (M1) for determining a
CC
CC toxicological response to an agent, which comprises comparing the
CC expression profile of one or more human toxic response genes to a
CC reference gene expression profile indicative of toxicity, and so
CC determining the presence of a toxic response to the agent. Also
CC described: (1) an array comprising one or more polynucleotides selected
CC from the genes corresponding to the partial sequences given in AB282842
CC to AB284764, or their fragments of at least 20 nucleotides, or
CC homologues; and (2) determining if a gene putatively identified to be a
CC toxic response gene plays a role on toxic response pathways by
CC determining the expression profile of the gene after exposure of cells
CC or a human subject to a known toxic pharmaceutical or industrial agent,
CC comprising: (a) exposing cells to an agent; (b) obtaining the test gene
CC expression profile for a putatively identified toxic response gene after
CC exposure to a known toxic pharmaceutical or industrial agent; and
CC (c) comparing the test profile to the expression profile of a gene with
CC a similar function or comparing the test profile to the expression
CC profile of that gene after exposure to other known toxic compounds. The
CC methods are useful for predicting and determining toxicological responses
CC on a cellular, organ or system level. The arrays comprising the human
CC genes are useful for toxicological screening of drugs, pharmaceutical
CC compounds and chemicals.
XX
XX Sequence 715 BP; 196 A; 173 C; 141 G; 205 T; 0 other;
SQ
Query Match 92.8%; Score 75.2; DB 25; Length 715;
Best Local Similarity 96.2%; Pred. No. 1e-16;
Matches 77; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 CCACATCAGACTGCTCAGCGCCACACATGTCCTGTCACACCCAGTGGCGGCTC 61
ABX47313
DB 49 CCTATCAGACTGCTCAGCGCCACACATGTCCTGTCACACCCAGTGGCGGCTC 108
QY 62 AGAAGGACAGATCCCGAC 81
DB 109 AAGAGGACAGATCCCGAC 128
RESULT 15
ID ABX47313 standard; cDNA; 434 BP.
XX
AC ABX47313;

XX 21-FEB-2003 (first entry)
DT
XX
DE Bovine EST associated with lactation/muscle/fat deposition #12478.
XX
KW Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
KW muscle deposition; fat deposition; genome mapping; gene identification;
KW gene analysis; cattle breeding.
XX
XX Bos Taurus.
OS
XX
XX US2002137139-A1.
PN
XX
XX 26-SEP-2002.
PD
XX
XX 24-SEP-2001; 2001US-0960352.
PF
XX
XX 12-JAN-1999; 99US-115707P.
PR
XX
XX 11-JAN-2000; 2000US-0480902.
PR
XX
XX (BYAT/) BYATT J C.
PA
XX (MATH/) MATHIALAGAN N.
PA (TAON/) TAO N.
PA (WARR/) WARREN W C.
XX
XX Byatt JC, Mathialagan N, Tao N, Warren WC;
PI
XX
XX WPI; 2003-110599/10.
DR
XX
XX New nucleic acid associated with lactation, and muscle and fat
PT
PT deposition, useful for genome mapping, gene identification and
PT analysis, cattle breeding, or for genetically improving cattle -
XX
XX Claim 2; SEQ ID NO 12478; 245pp; English.
PS
XX
XX The invention relates to a purified nucleic acid molecule associated with
CC lactation or muscle and fat deposition (designated LMFD), derived
CC from cattle, and the LMFD nucleic acid can specifically hybridise to a
CC second nucleic acid molecule comprising any of 15112 nucleotide
CC sequences, appearing as ABX34836-ABX49947, or complements of them.
CC Also included are; (1) a transformed cell having a nucleic acid
CC comprising an LMFD nucleic acid linked to a promoter and a 3' non-
CC translated sequence and addition of polyadenylated ribonucleotides to a 3' end
CC transcription and addition of polyadenylated ribonucleotides to a 3' end
CC of the mRNA molecule; and (2) determining a level or pattern of a
CC molecule in a bovine cell or tissue comprising: (a) incubating a marker
CC nucleic acid (comprising any of the 15112 nucleic acid sequences or its
CC complement or fragment) with a complementary nucleic acid molecule
CC obtained from the bovine cell or tissue, where hybridisation between the
CC marker nucleic acid and the complementary nucleic acid permits the
CC detection of the molecule; and (b) detecting the level or pattern of the
CC complementary nucleic acid, where the detection of the complementary
CC nucleic acid is predictive of the level or pattern of the molecule.
CC The LMFD nucleic acid is used for determining a level or pattern
CC of a molecule in a bovine cell or tissue. It is useful for genome
CC mapping, gene identification and analysis, cattle breeding, preparation
CC of constructs for use in cattle gene expression, or for genetically
CC improving cattle. The present sequence is one of the 15112 bovine
CC LMFD EST (expressed sequence tag) nucleic acids.
CC Note: the present sequence was not shown in the specification but
CC was obtained in electronic format from the USPTO web site:
CC seqdata.uspto.gov/sequence.html?docID=20020137139.
XX
XX Sequence 434 BP; 135 A; 109 C; 86 G; 104 T; 0 other;
SQ
Query Match 90.1%; Score 73; DB 25; Length 434;
Best Local Similarity 93.8%; Pred. No. 5.4e-16;
Matches 76; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 GCCACATCAGACTGGCTCAACGCCAACCAATGTCCTGTCACACCCAGTGGCGGCTC 60
ABX47313
DB 36 GCCACATCAGACTGGCTCAATGCCACACATGTACCTGTCACCCAGTGGCGGCTC 95

Db 4370 CAAGAGGACAGATCCAGC 4390

RESULT 12

AC89480

ID AAC89480 standard; DNA; 5761 BP.

XX AC AAC89480;

XX DT 13-MAR-2001 (first entry)

XX DE Human TL405 carbamyl phosphate synthase I DNA.

XX KW Urea cycle; carbamyl phosphate synthase I; CPSI; hepatitis;

XX KW sclerosis; pulmonary hypertension; bone marrow; gene therapy; ds.

XX OS Homo sapiens.

XX PN WO200073322-A1.

XX PD 07-DEC-2000.

XX PF 01-JUN-2000; 2000WO-US15079.

XX PR 01-JUN-1999; 99US-0323472.

XX PA (UYVA-) UNIV VANDERBILT.

XX PI Summar ML, Christman BW;

XX FI WPI; 2001-049926/06.

XX DR Detecting susceptibility to suboptimal urea cycle function, e.g. bone marrow transplant toxicity, comprises identifying a polymorphism in the gene for carbamylphosphate synthase

XX PS Disclosure; Page 134-140; 171pp; English.

XX CC The present invention relates to screening for susceptibility to suboptimal urea cycle function or to bone marrow transplant toxicity by detecting a polymorphism in the carbamyl phosphate synthase I (CPSI) gene. The method is used to detect subjects at risk of hepatitis, sclerosis, pulmonary hypertension and bone marrow transplant toxicity, these conditions can be treated or prevented by administration of a nitric oxide precursor or by gene therapy (administration of sequences that encode CPSI).

XX SQ Sequence 5761 BP; 1624 A; 1210 C; 1309 G; 1618 T; 0 other;

Query Match 98.0%; Score 79.4; DB 22; Length 5761;

Best Local Similarity 98.8%; Pred. No. 5,4e-18;

Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTCCACCCAGTGGCATGGCGTCT 60

Db 4300 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTCCACCCAGTGGCATGGCGTCT 4359

Qy 61 CAAGAAGGACAGATCCAGC 81

Db 4360 CAAGAAGGACAGATCCAGC 4380

RESULT 13

ABZ35442

ID ABZ35442 standard; cDNA; 5772 BP.

XX AC ABZ35442;

XX DT 05-FEB-2003 (first entry)

XX DE Human gene expression profile polynucleotide SEQ ID NO 553.

XX KW Human; artery; endothelium; umbilical; vein; aorta; pulmonary artery;

KW bronchial epithelium; prostate; muscle; lung fibroblast; osteoblast; tumour; microarray; genome mapping; antibiotic; antiviral; antifungal; gene expression; gene; ss.

XX OS Homo sapiens.

XX PN WO200274979-A2.

XX PD 26-SEP-2002.

XX PF 20-MAR-2002; 2002WO-US08456.

XX PR 20-MAR-2001; 2001US-276947P.

XX PA (ORTH) ORTHO CLINICAL DIAGNOSTICS INC.

XX PI Wan J, Wang Y;

XX FI WPI; 2002-740862/80.

XX DR New gene expression profile generated from primary, endothelial, epithelial, and muscle cell types, useful for identifying disease pathologies involving alterations of gene expression, e.g. cancer

XX PS Disclosure; Page 691-693; 850pp; English.

XX CC The invention relates to a gene expression profile comprising one or more genes (ABZ34889-ABZ35692) and generated from a cell type. The cell type is a coronary artery endothelium, umbilical artery or vein endothelium, aortic endothelium, dermal microvascular endothelium, pulmonary artery endothelium, myometrium microvascular endothelium, keratinocyte epithelium, bronchial epithelium, mammary epithelium, prostate epithelium, renal cortical epithelium, renal proximal tubule epithelium, small airway epithelium, renal epithelium, umbilical artery smooth muscle, neonatal dermal fibroblast, pulmonary artery smooth muscle, dermal fibroblast, neural progenitor cells, skeletal muscle, astrocytes, aortic smooth muscle, mesangial cells, coronary artery smooth muscle, bronchial smooth muscle, uterine smooth muscle, lung fibroblast, osteoblasts or prostate stromal cell. The gene expression profile is used for determining the level of RNA expression for a sample, determining the phenotype of a cell and distinguishing cell types. The gene or a protein expression profile is useful in identifying disease pathologies involving alterations of gene expression. The assessment of expression CC profiles may provide meaningful information with respect to tumour type and stage, treatment methods, and prognosis. The gene or protein expression profile may also be used for creating microarrays. The microarray is useful for genetic and physical mapping of genomes, DNA sequencing, genetic or medical diagnosis, genotyping of organisms, confirming cell or tissue identifications and in identifying promising antibiotics, antiviral or antifungal agents.

XX SQ Sequence 5772 BP; 1602 A; 1180 C; 1295 G; 1571 T; 124 other;

Query Match 98.0%; Score 79.4; DB 24; Length 5772;

Best Local Similarity 98.8%; Pred. No. 5,4e-18;

Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTCCACCCAGTGGCATGGCGTCT 60

Db 4310 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTCCACCCAGTGGCATGGCGTCT 4369

Qy 61 CAAGAAGGACAGATCCAGC 81

Db 4370 CAAGAAGGACAGATCCAGC 4390

RESULT 14

ABZ84741

ID ABZ84741 standard; cDNA; 715 BP.

XX AC ABZ84741;

XX DT 14-MAY-2003 (first entry)

```
Db 157 GCCACATCAGACTGGCTCAAGCCCAACAATGTCCCTGCCACCCAGTGGCATGGCGGTCT 216
QY 61 CAAGAAGGACAGATCCCGAGC 81
    |||||||
Db 217 CAAGAAGGACAGATCCCGAGC 237

RESULT 10
ABX939366
ID ABX939366 standard; cDNA; 531 BP.
XX
AC ABX939366;
XX
DT 22-MAY-2003 (first entry)
XX
DE Lung cancer therapy and diagnosis associated cDNA #93.
XX
KW Lung cancer; cytostatic; vaccine; gene therapy; cancer;
KW gene; ss.
XX
OS Homo sapiens.
XX
PN US2002172952-A1.
XX
PD 21-NOV-2002.
XX
PF 10-JUL-2001; 2001US-0902941.
XX
PR 30-JUN-1999; 99US-0346492.
PR 15-OCT-1999; 99US-0419356.
PR 17-DEC-1999; 99US-0466867.
PR 30-DEC-1999; 99US-0476300.
PR 06-MAR-2000; 2000US-0519642.
PR 22-MAR-2000; 2000US-0533077.
PR 10-APR-2000; 2000US-0546259.
PR 27-APR-2000; 2000US-0560406.
PR 05-JUN-2000; 2000US-0589184.
PR 11-JUL-2000; 2000US-0614124.
PR 29-AUG-2000; 2000US-0651563.
PR 08-SEP-2000; 2000US-0658824.
PR 26-SEP-2000; 2000US-0671325.
PR 06-OCT-2000; 2000US-0677419.
PR 30-OCT-2000; 2000US-0702705.
PR 13-DEC-2000; 2000US-0736457.
PR 03-MAY-2001; 2001US-0849626.
XX
PA (CORI-) CORIXA CORP.
XX
PI Henderson RA, Wang T, Watanabe Y, Johnson JC, Retter MW, Durham M;
PI Carter D, Fanger GR, Vedvick TS, Bangur CS, McNabb A;
XX
DR WPI: 2003-328427/31.
XX
PT New polynucleotide, useful for preparing a composition for treating or
PT inhibiting development of cancer, e.g. lung cancer.
XX
PS Example 1; SEQ ID NO 93; 82pp; English.
XX
CC The invention describes an isolated polynucleotide comprising one of 32
CC sequences, complement or degenerate variants of them. The polynucleotide
CC is useful for preparing a composition e.g. a vaccine or for gene therapy,
CC for treating or inhibiting development of cancer, e.g. lung cancer.
CC This sequence represents a polynucleotide associated with the
CC compositions and methods for the therapy and diagnosis of lung cancer.
XX
SQ Sequence 531 BP; 157 A; 139 C; 108 G; 127 T; 0 other;

Query Match 98.0%; Score 79.4; DB 25; Length 531;
Best Local Similarity 98.8%; Pred. No. 3.2e-18;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAAGCCCAACAATGTCCCTGCCACCCAGTGGCATGGCGGTCT 60
    |||||||
```

```
Db 157 GCCACATCAGACTGGCTCAAGCCCAACAATGTCCCTGCCACCCAGTGGCATGGCGGTCT 216
QY 61 CAAGAAGGACAGATCCCGAGC 81
    |||||||
Db 217 CAAGAAGGACAGATCCCGAGC 237

RESULT 11
AAH57467
ID AAH57467 standard; cDNA; 5195 BP.
XX
AC AAH57467;
XX
DT 10-SEP-2001 (first entry)
XX
DE Human liver cell specific cDNA sequence SEQ ID NO:307.
XX
KW Human; tissue specific; diagnosis; brain; heart; skeletal muscle;
KW lung; liver; uterus; ovary; stomach; intestine; kidney; pancreas; ss;
KW metabolic disease; developmental disease; cytostatic; immunomodulatory;
KW neuroprotective; gene therapy; cancer; immunopathology; neuropathology.
XX
OS Homo sapiens.
XX
PN WO200132927-A2.
XX
PD 10-MAY-2001.
XX
PF 02-NOV-2000; 2000WO-US30396.
XX
PR 04-NOV-1999; 99US-0163508.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Sornasse T, Seilhamer JJ, Watson GA;
XX
DR WPI: 2001-291057/30.
XX
PT New cell and tissue specific polynucleotides useful for diagnosis,
PT prognosis or monitoring of treatments for disorders where the gene is
PT associated with a cancer, immunopathology or neuropathology.
XX
PS Claim 1; Page 230-231; 327pp; English.
XX
CC AAH57161 to AAH57576 represent cell and tissue specific polynucleotide
CC sequences (I). (I) can have cytostatic, immunomodulatory and
CC neuroprotective activities, and can be used in gene therapy. (I) and
CC proteins (II) encoded by them are used in high throughput screening
CC assays to select DNA molecules, RNA molecules, peptide nucleic acids,
CC mimetics, peptides, proteins, agonists, antagonists, antibodies or
CC their fragments, immunoglobulins, inhibitors, drug compounds and
CC pharmaceutical agents. Expression of (I) in a sample indicates the
CC differentiation of embryonic stem cells into a tissue selected from
CC brain, heart, kidney, liver, lung, skeletal muscle or pancreatic
CC tissues. (I) and (II) are used to produce an expression profile that
CC defines a metabolic or developmental process, treatment, condition,
CC disease or disorder. The gene profile can be used for diagnosis,
CC prognosis or monitoring of treatments and for investigating a
CC predisposition to a disorder where the gene is associated with a
CC cancer, immunopathology or neuropathology.
XX
SQ Sequence 5195 BP; 1469 A; 1102 C; 1211 G; 1413 T; 0 other;

Query Match 98.0%; Score 79.4; DB 22; Length 5195;
Best Local Similarity 98.8%; Pred. No. 5.3e-18;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAAGCCCAACAATGTCCCTGCCACCCAGTGGCATGGCGGTCT 60
    |||||||
Db 4310 GCCACATCAGACTGGCTCAAGCCCAACAATGTCCCTGCCACCCAGTGGCATGGCGGTCT 4369
    |||||||

QY 61 CAAGAAGGACAGATCCCGAGC 81
    |||||||
```



```
Query Match      100.0%; Score 81; DB 22; Length 5761;
Best Local Similarity 100.0%; Pred. No. 1.5e-18;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACATCAGCTGCTCAAGGCCACACAAATGTCCTGCCAACCCAGTGGCATGGCGTCT 60
    |||||||
Db 4300 GCCACATCAGCTGCTCAAGGCCACACAAATGTCCTGCCAACCCAGTGGCATGGCGTCT 4359

QY 61 CAAGAAGGACAGAAATCCCGAGC 81
    |||||||
Db 4360 CAAGAAGGACAGAAATCCCGAGC 4380

RESULT 4
AAC89487
ID AAC89487 standard; DNA; 5761 BP.
XX
AC AAC89487;
XX
DT 13-MAR-2001 (first entry)
XX
DE Human carbamyl phosphate synthase DNA.
XX
KW Urea cycle; carbamyl phosphate synthase I; CPSI; hepatitis;
XX sclerosis; pulmonary hypertension; bone marrow; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200073322-A1.
XX
PD 07-DEC-2000.
XX
PF 01-JUN-2000; 2000WO-US15079.
XX
PR 01-JUN-1999; 99US-0323472.
XX
PA (UYVA-) UNIV VANDERBILT.
XX
PI Sumnar ML, Christman BW;
XX WPI; 2001-049926/06.
XX
PT Detecting susceptibility to suboptimal urea cycle function, e.g. bone
PT marrow transplant toxicity, comprises identifying a polymorphism in the
PT gene for carbamylphosphate synthase -
XX
PS Claim 65; Page 146; 171pp; English.
XX
CC The present invention relates to screening for susceptibility to
CC suboptimal urea cycle function or to bone marrow transplant toxicity
CC by detecting a polymorphism in the carbamyl phosphate synthase I
CC (CPSI) gene. The method is used to detect subjects at risk of
CC hepatitis, sclerosis, pulmonary hypertension and bone marrow
CC transplant toxicity. These conditions can be treated or prevented
CC by administration of a nitric oxide precursor or by gene
CC therapy (administration of sequences that encode CPSI).
XX
SQ Sequence 5761 BP; 1625 A; 1209 C; 1309 G; 1618 T; 0 other;

Query Match      100.0%; Score 81; DB 22; Length 5761;
Best Local Similarity 100.0%; Pred. No. 1.5e-18;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACATCAGCTGCTCAAGGCCACACAAATGTCCTGCCAACCCAGTGGCATGGCGTCT 60
    |||||||
Db 4300 GCCACATCAGCTGCTCAAGGCCACACAAATGTCCTGCCAACCCAGTGGCATGGCGTCT 4359

QY 61 CAAGAAGGACAGAAATCCCGAGC 81
    |||||||
Db 4360 CAAGAAGGACAGAAATCCCGAGC 4380

RESULT 5
AAC89481
ID AAC89481 standard; DNA; 5761 BP.
XX
AC AAC89481;
XX
DT 13-MAR-2001 (first entry)
XX
DE Target 5' and 3' sequence.
XX
KW Urea cycle; carbamyl phosphate synthase I; CPSI; hepatitis;
XX sclerosis; pulmonary hypertension; bone marrow; gene therapy; ds.
XX
OS Homo sapiens.
XX
```

```
AAC89488
ID AAC89488 standard; DNA; 5761 BP.
XX
AC AAC89488;
XX
DT 13-MAR-2001 (first entry)
XX
DE Human carbamyl phosphate synthase DNA.
XX
KW Urea cycle; carbamyl phosphate synthase I; CPSI; hepatitis;
XX sclerosis; pulmonary hypertension; bone marrow; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200073322-A1.
XX
PD 07-DEC-2000.
XX
PF 01-JUN-2000; 2000WO-US15079.
XX
PR 01-JUN-1999; 99US-0323472.
XX
PA (UYVA-) UNIV VANDERBILT.
XX
PI Sumnar ML, Christman BW;
XX WPI; 2001-049926/06.
XX
PT Detecting susceptibility to suboptimal urea cycle function, e.g. bone
PT marrow transplant toxicity, comprises identifying a polymorphism in the
PT gene for carbamylphosphate synthase -
XX
PS Claim 65; Page 156-161; 171pp; English.
XX
CC The present invention relates to screening for susceptibility to
CC suboptimal urea cycle function or to bone marrow transplant toxicity
CC by detecting a polymorphism in the carbamyl phosphate synthase I
CC (CPSI) gene. The method is used to detect subjects at risk of
CC hepatitis, sclerosis, pulmonary hypertension and bone marrow
CC transplant toxicity. These conditions can be treated or prevented
CC by administration of a nitric oxide precursor or by gene
CC therapy (administration of sequences that encode CPSI).
XX
SQ Sequence 5761 BP; 1626 A; 1209 C; 1308 G; 1618 T; 0 other;

Query Match      100.0%; Score 81; DB 22; Length 5761;
Best Local Similarity 100.0%; Pred. No. 1.5e-18;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACATCAGCTGCTCAAGGCCACACAAATGTCCTGCCAACCCAGTGGCATGGCGTCT 60
    |||||||
Db 4300 GCCACATCAGCTGCTCAAGGCCACACAAATGTCCTGCCAACCCAGTGGCATGGCGTCT 4359

QY 61 CAAGAAGGACAGAAATCCCGAGC 81
    |||||||
Db 4360 CAAGAAGGACAGAAATCCCGAGC 4380

RESULT 6
AAC89481
ID AAC89481 standard; DNA; 495 BP.
XX
AC AAC89481;
XX
DT 13-MAR-2001 (first entry)
XX
DE Target 5' and 3' sequence.
XX
KW Urea cycle; carbamyl phosphate synthase I; CPSI; hepatitis;
XX sclerosis; pulmonary hypertension; bone marrow; gene therapy; ds.
XX
OS Homo sapiens.
XX
```


DR WPI; 1993-006237/01.
DR P-PSDB; AAR30636.
XX Carbamyl phosphate synthetase I gene - used to detect diseases
PT caused by carbamyl phosphatase synthetase by using overlapped
PT cDNA contg. total human cDNA sequence
XX
PS Claim 1; Page 6-8; 12pp; Japanese.
XX This sequence is the human carbamyl phosphate synthetase I (CPSI)
CC gene. This sequence or fragments of it may be used as probes to
CC detect mutations in the CPSI gene. CPSI deficiency disease may be
CC detected by using overlapping cDNA representing the full length cDNA
CC sequence of human CPSI. This sequence was isolated using three
CC amplified fragments derived from protein coding regions of the rat
CC CPSI cDNA as probes.
XX
SQ Sequence 5215 BP; 1484 A; 1104 C; 1205 G; 1422 T; 0 other;
Query Match 100.0%; Score 81; DB 14; Length 5215;
Best Local Similarity 100.0%; Pred. No. 1.4e-18;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCCACATCAGACTGGCTCAAGCCCAACAATGTCCTGCCAACCAGTGGCGCTCT 60
|||||
Db 4295 GCCACATCAGACTGGCTCAAGCCCAACAATGTCCTGCCAACCAGTGGCGCTCT 4354
QY 61 CAAGAAGCAGAGATCCCGC 81
|||||
Db 4355 CAAGAAGCAGAGATCCCGC 4375
RESULT 2
ABN95271
ID ABN95271 standard; DNA; 5215 BP.
XX
AC ABN95271;
XX
DT 13-AUG-2002 (first entry)
XX
DE Gene #1769 used to diagnose liver cancer.
XX
KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW metastatic liver tumor; cytostatic; expression profile; disease state;
KW disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
OS Homo sapiens.
XX
PN WO200229103-A2.
XX
PD 11-APR-2002.
XX
PF 02-OCT-2001; 2001WO-US30589.
XX
PR 02-OCT-2000; 2000US-237054P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX
DR WPI; 2002-426119/45.
XX
PT Diagnosing and detecting the progression of liver cancer,
PT hepatocellular carcinoma or metastatic liver tumor in a patient,
PT involves detecting the level of expression of two or more genes in a
PT liver tissue sample -
XX
PS Claim 1; SEQ ID NO 1769; 298pp; English.
XX
CC The invention relates to a novel method for diagnosing and detecting the
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC tumor in a patient, and differentiating metastatic liver cancer from
CC hepatocellular carcinoma in a patient, involving detecting the level of

CC expression of two or more genes represented in ABN93503-ABN97455 in a
CC tissue sample. The method of the invention has hepatotropic, and
CC cytostatic activity. The method is useful for diagnosing and detecting
CC the progression of liver cancer, hepatocellular carcinoma and metastatic
CC liver carcinoma in a patient. The method is useful for identifying
CC expression profiles which serve as useful diagnostic markers as well as
CC markers that can be used to monitor disease states, disease progression,
CC drug toxicity, drug efficacy and drug metabolism.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 5215 BP; 1484 A; 1102 C; 1207 G; 1422 T; 0 other;
Query Match 100.0%; Score 81; DB 24; Length 5215;
Best Local Similarity 100.0%; Pred. No. 1.4e-18;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCCACATCAGACTGGCTCAAGCCCAACAATGTCCTGCCAACCAGTGGCGCTCT 60
|||||
Db 4295 GCCACATCAGACTGGCTCAAGCCCAACAATGTCCTGCCAACCAGTGGCGCTCT 4354
QY 61 CAAGAAGCAGAGATCCCGC 81
|||||
Db 4355 CAAGAAGCAGAGATCCCGC 4375
RESULT 3
AAC89479
ID AAC89479 standard; DNA; 5761 BP.
XX
AC AAC89479;
XX
DT 13-MAR-2001 (first entry)
XX
DE Human N1405 carbamyl phosphate synthase I DNA.
XX
KW Urea cycle; carbamyl phosphate synthase I; CPSI; hepatitis;
KW sclerosis; pulmonary hypertension; bone marrow; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200073322-A1.
XX
PD 07-DEC-2000.
XX
PF 01-JUN-2000; 2000WO-US15079.
XX
PR 01-JUN-1999; 99US-0323472.
XX
PA (UYVA-) UNIV VANDERBILT.
XX
PI Summar ML, Christman BW;
XX
DR WPI; 2001-049926/06.
XX
PT Detecting susceptibility to suboptimal urea cycle function, e.g. bone
PT marrow transplant toxicity, comprises identifying a polymorphism in the
PT gene for carbamylphosphate synthase -
XX
PS Disclosure; Page 124-130; 171pp; English.
XX
CC The present invention relates to screening for susceptibility to
CC suboptimal urea cycle function or to bone marrow transplant toxicity
CC by detecting a polymorphism in the carbamyl phosphate synthase I
CC (CPSI) gene. The method is used to detect subjects at risk of
CC hepatitis, sclerosis, pulmonary hypertension and bone marrow
CC transplant toxicity. These conditions can be treated or prevented
CC by administration of a nitric oxide precursor or by gene
CC therapy (administration of sequences that encode CPSI).
XX
SQ Sequence 5761 BP; 1625 A; 1209 C; 1309 G; 1618 T; 0 other;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2003, 22:19:36 ; Search time 201.5 Seconds
(without alignments)
1085.134 Million cell updates/sec

Title: US-09-585-077c-1_COPY_4300_4380

Perfect score: 81

Sequence: 1 gccacatcagactggtctaa.....aagaagacgaatccaccg 81

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_19Jun03.*

```

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					Description
Result No.	Score	Query Match	Length	ID	
1	81	100.0	5215	14	AAQ34768
2	81	100.0	5215	24	ABN95271
3	81	100.0	5761	22	AAC89479
4	81	100.0	5761	22	AAC89487
5	81	100.0	5761	22	AAC89488
6	79.4	98.0	495	22	AAC89481
7	79.4	98.0	531	22	AAF68175
8	79.4	98.0	531	24	ABK38086
hCPSI gene. Homo					
Gene #1769 used to					
Human N1405 carbam					
Human carbamyl pho					
Human carbamyl pho					
Target 5' and 3' s					
Human lung tumour					
CDNA encoding clon					

9	79.4	98.0	531	25	ACA10415	Human lung cancer-
10	79.4	98.0	531	25	ABX9366	Lung cancer therap
11	79.4	98.0	5195	22	AAH57467	Human liver cell s
12	79.4	98.0	5761	22	AAC89480	Human N1405 carbam
13	79.4	98.0	5772	24	ABZ35442	Human gene express
14	75.2	92.8	715	25	ABZ84741	Toxicologically re
15	73	90.1	434	25	ABX47313	Bovine EST associa
16	69	85.2	816	24	ABA93013	Human cDNA clone B
17	62.8	77.5	459	22	AAH57259	Human liver specif
18	46	56.8	580	24	ABT08945	Phase-1 Rat CT gen
19	40.2	49.6	242	25	ABX36653	Bovine EST associa
20	31	38.3	31	22	AAI31280	Human single nucle
21	29.4	36.3	3591	22	AAI02819	Human reproductive
22	27.4	33.8	646	24	ABT09297	Phase-1 Rat CT gen
23	26.8	33.1	8098	22	AAK90713	Human digestive sy
24	26.8	33.1	9556	22	AAK90714	Human digestive sy
25	26.6	32.8	2372	8	AAH70814	Aspergillus niger
26	26.4	32.6	2153	24	ABT19980	Mouse ischaemic co
27	26.4	32.6	2168	18	AAV00499	Murine alpha-(2) s
28	26.2	32.3	6564	22	AAF31637	Mycobacterium tube
29	26.2	32.3	4403765	22	AAI99683	Mycobacterium tube
30	26.2	32.3	4411529	22	AAI99682	Mycobacterium tube
31	26	32.1	375	22	AAH36700	Human colon cancer
32	26	32.1	439	22	AAI88303	Human polynucleoti
33	26	32.1	987	21	AAC79247	Human lung tumour-
34	26	32.1	987	23	AAD23323	Human lung tumour-
35	26	32.1	1987	20	AAZ53006	Human prostate tum
36	26	32.1	2236	20	AAZ41294	Human normal ovari
37	25.8	31.9	349980	22	AAH41225	Pyrococcus abyssi
38	25.8	31.9	349980	22	AAH41226	Pyrococcus abyssi
39	25.6	31.6	799	22	AAI93930	Human neuroblastom
40	25.6	31.6	1036	24	ABK48308	Human inhibitory G
41	25.6	31.6	1036	25	ACA56710	Human signalling p
42	25.6	31.6	1202	19	AAV71981	A nucleotide seque
43	25.4	31.4	288	25	ABX54929	Bovine EST associa
44	25.4	31.4	26225	22	AAZ35776	Human cardiovascular
45	25	30.9	25	25	ABZ84401	Toxicologically re

ALIGNMENTS

RESULT 1
AAQ34768
ID AAQ34768 standard; DNA; 5215 BP.
XX AC AAQ34768;
XX AC
DT 06-MAY-1993 (first entry)
XX hCPSI gene.
XX hCPSI gene.
KW Human; carbamyl phosphate synthetase I; CPSI; probe; mutation;
KW CPSI deficiency disease; rat; ss.
XX Homo sapiens.
XX Homo sapiens.
FH Key Location/Qualifiers
CDS 119..4621
FT /*tag= a
XX
XX JP04335889-A.
XX
PD 24-NOV-1992.
XX
XX 09-MAY-1991; 91JP-0135902.
XX
PR 09-MAY-1991; 91JP-0135902.
XX
XX (HARA/) HARAGUCHI Y.
PA (MATS/) MATSUDA I.
PA (MORI/) MORI M.

BASE COUNT 1266 a 967 c 1113 g 1157 t
ORIGIN
Query Match 98.0%; Score 79.4; DB 9; Length 4503;
Best Local Similarity 98.8%; Pred. No. 3.9e-17;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GCCACATCAGACTGGCTCAACGCCAACCAATGCTCCCTGCCAACCCAGTGGCATGGCGCTCT 60
|||||
Db 4177 GCCACATCAGACTGGCTCAACGCCAACCAATGCTCCCTGCCAACCCAGTGGCATGGCGCTCT 4236
QY 61 CAAGAAGGACAGATCCCGC 81
|||||
Db 4237 CAAGAAGGACAGATCCCGC 4257
RESULT 14
AR184417 5761 bp DNA linear PAT 20-APR-2002
LOCUS AR184417
DEFINITION Sequence 3 from patent US 6346382.
ACCESSION AR184417
VERSION AR184417.1 GI:20230382
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
AUTHORS Summar,M.I. and Christman,B.W.
TITLE Human carbamyl phosphate synthetase I polymorphism and diagnostic methods related thereto
JOURNAL Patent: US 6346382-A 3 12-FEB-2002;
FEATURES
source Location/Qualifiers
1..5761 /organism="unknown"
BASE COUNT 1624 a 1210 c 1309 g 1618 t
ORIGIN
Query Match 98.0%; Score 79.4; DB 6; Length 5761;
Best Local Similarity 98.8%; Pred. No. 3.8e-17;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GCCACATCAGACTGGCTCAACGCCAACCAATGCTCCCTGCCAACCCAGTGGCATGGCGCTCT 60
|||||
Db 4300 GCCACATCAGACTGGCTCAACGCCAACCAATGCTCCCTGCCAACCCAGTGGCATGGCGCTCT 4359
QY 61 CAAGAAGGACAGATCCCGC 81
|||||
Db 4360 CAAGAAGGACAGATCCCGC 4380
RESULT 15
AF154830 5761 bp mRNA linear PRI 09-JUN-1999
LOCUS AF154830
DEFINITION Homo sapiens carbamyl phosphate synthetase I mRNA, complete cds.
ACCESSION AF154830
VERSION AF154830.1 GI:5020419
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 5761)
Summar,M.
Direct Submission
Submitted (27-MAY-1999), Pediatrics, Vanderbilt University Medical
Center, D2205 Medical Center North, Nashville, TN 37232, USA
FEATURES
source Location/Qualifiers
1..5761 /organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="2"

/map="2q35"
124..4626
/note="CPSI"
/codon_start=1
/product="carbamyl phosphate synthetase I"
/protein_id="AAD38072.1"
/gb_xref="GI:5020420"
/translation="MTRILTAFAKVVRLTKTGFNTVAHQKWKFSRPGIRLLSVKQAQ
TAHVLDTGTMKGYSPGHSPVAVGVNTGLGYPEALTDPAKGOILTMANPIIG
NGADPTALDELGSYLENGIKVSGLLVDYSDYNHMLATKSLGOWLQEKVPA
IYGVTRMTXIIEDKTMGLKEEGOPVDVDPNKONLIAEVSTKDVYVKGNT
KYAVDCGKNNVIRLLVYKRAEYHLVFNWDFTKMEYDGLIAGGPNALAPLQ
NVKRIEEDREPLFGISTONLITGLAAGATYKMSNARGQNPQVINTINKQAFITA
QNHGVALDNTLPAGKFLFVYNNDTNEGIMHESKFFFAVQHPHVPITPGDITDLYFD
SFFSLIKKATITISVLPKPAVASRVEVSKVILGSGLSGAGGFDSGQAVK
AMKEENVKTVLMNPNIASQTVNEVGLKQADTVYFLPTPQVTEVTKAEQDGLILGM
GGOTALNGVELFKRGVLYEYGVKLTGSVIMATEDROLFSDKLNEINERKIAPSA
VESIEDALKAADTIGYPMIRSAVALGGISGICPNRETIMDLSTKAFAMTNOILVEK
SVTGHKEIYEVYDADDNCVVCNMENVDAMGVHTGDSVVVAPACTLSNAEFOMLR
TSINVRHLGIVGCNIOFALHPTSMEXCIIEVNRSLRSALASRATGIPPLAFAK
IALGPLPEIKNVYSGKTSACFEPSLDYXMTKIPRWOLDRFHGTSSRIGSSKSVGEV
MAIGTFEESFKALRMCHPSIBGFTPLPNKEMPNDLDELKELSEPSSTRIAYIAK
AIDNMSLDEIEKLYIDKWFLYKMRDILNKEKTLKGINSESTETLTKRAKEIGFSD
KOISCLIGTEAOTRELRLKKNHPWKQIDTLAAEYPSVTNYLYTVYNGQSHDNFD
DHGMVLCGPYHIGSVFEDWCAYSSIRTLROLGKKTVVVNCNPTVTSDFDCDKL
YPEELSLERILDIYHOEACGGCLISVGGIENLAVPLXKNGVKIMGTSPLOIDRAED
RSIFSAYLDELKVAQAPKAVNTLNEALEFAKSDVDYDPCLLRPSYVLSGSMNVFSED
EMKKFLEATRVSOEHFVVLITKEVGEAREVEDAVGDKGRVISHALSEVEDVGHSG
DATLMLPTQTSQGAIEKVKDATRKAKAFISGFNFQVFLKGNVDVLYIECNLRASR
SFPFKTIGVDVIDVATKVMIGENVDEKHLFDHPFIIPADYVAIKAPMFWPRLRD
ADPILRCMASTGEVACFGEIGIHTAPLKMSTGFKIPQKILIGIQQSFRRPRTGVA
EOLHNEGPKLRATEATSDMLNANNVATPVAMPQSQGNPSSLSSIRKLJRDGSLDLVI
NLPNNNTKEVHDNYVIRTAVDGSIPLLTNFQVTKLFAEAVQKSRKVDKSLFHYRQY
SAGKAA"

BASE COUNT 1624 a 1210 c 1309 g 1618 t
ORIGIN
Query Match 98.0%; Score 79.4; DB 9; Length 5761;
Best Local Similarity 98.8%; Pred. No. 3.8e-17;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GCCACATCAGACTGGCTCAACGCCAACCAATGCTCCCTGCCAACCCAGTGGCATGGCGCTCT 60
|||||
Db 4300 GCCACATCAGACTGGCTCAACGCCAACCAATGCTCCCTGCCAACCCAGTGGCATGGCGCTCT 4359
QY 61 CAAGAAGGACAGATCCCGC 81
|||||
Db 4360 CAAGAAGGACAGATCCCGC 4380

Search completed: September 12, 2003, 23:21:50
Job time : 1423.5 secs

```

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 157 a 139 c 108 g 127 t
ORIGIN
Query Match 98.0%; Score 79.4; DB 6; Length 531;
Best Local Similarity 98.8%; Pred. No. 4.7e-17;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGCTCAAGCCCAACAATGTCCTGCGCAACCCAGTGGCATGGCGGCTT 60
|||||
Db 157 GCCACATCAGACTGCTCAAGCCCAACAATGTCCTGCGCAACCCAGTGGCATGGCGGCTT 216
|||||

QY 61 CAAGAAGGACAGATCCCGAGC 81
|||||
Db 217 CAAGAAGGACAGATCCCGAGC 237

RESULT 12
AX367383
LOCUS AX367383 531 bp DNA linear PAT 17-FEB-2002
DEFINITION Sequence 93 from Patent WO0204514.
ACCESSION AX367383
VERSION AX367383.1 GI:18855487
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Wang,T., Watanabe,Y., Henderson,R.A., Johnson,J.C., Retter,M.W.,
Marnerakis,M., Carter,D., Fanger,G.R., Vedvick,T.S., Bangur,C.S.,
McNabb,A., Fanger,N., Switzer,A., McNeill,P.D. and Clapper,J.D.
COMPOSITIONS and methods for the therapy and diagnosis of lung
cancer
JOURNAL Patent: WO 0204514-A 93 17-JAN-2002;
CORIAX CORPORATION (US)
FEATURES
Location/Qualifiers
1..531
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 157 a 139 c 108 g 127 t
ORIGIN
Query Match 98.0%; Score 79.4; DB 6; Length 531;
Best Local Similarity 98.8%; Pred. No. 4.7e-17;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGCTCAAGCCCAACAATGTCCTGCGCAACCCAGTGGCATGGCGGCTT 60
|||||
Db 157 GCCACATCAGACTGCTCAAGCCCAACAATGTCCTGCGCAACCCAGTGGCATGGCGGCTT 216
|||||

QY 61 CAAGAAGGACAGATCCCGAGC 81
|||||
Db 217 CAAGAAGGACAGATCCCGAGC 237

RESULT 13
HSCPSIMUT HSCPSIMUT 4503 bp mRNA linear PRI 24-SEP-1998
LOCUS Homo sapiens mRNA for carbamyl phosphate synthetase I.
DEFINITION
ACCESSION Y15793
VERSION Y15793.1 GI:3228247
KEYWORDS carbamyl-phosphate synthetase I; CPSI gene; mutation.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Finckh,U., Kohlschutter,A., Schafer,H., Sperhake,K., Colombo,J.P.
and Gal,A.

```

```

TITLE Prenatal diagnosis of carbamoyl phosphate synthetase I deficiency
JOURNAL Hum. Mutat. 12 (3), 206-211 (1998)
MEDLINE 98375696
PUBMED 9711878
AUTHORS Finckh,U.
DIRECT SUBMISSION
TITLE Direct Submission
JOURNAL Submitted (05-DEC-1997) U. Finckh, Dept. of Human Genetics,
University Hospital Eppendorf, Butenfeld 42, 22529 Hamburg, FRG
REMARK revised by [3]
REFERENCE 3 (bases 1 to 4503)
AUTHORS Finckh,U.
DIRECT SUBMISSION
TITLE Direct Submission
JOURNAL Submitted (12-JUN-1998) U. Finckh, Dept. of Human Genetics,
University Hospital Eppendorf, Butenfeld 42, 22529 Hamburg, FRG
COMMENT On Jun 16, 1998 this sequence version replaced gi:2959899.
Related sequence D90282.
FEATURES
Location/Qualifiers
1..4503
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="2"
/map="q35"
/tissue_type="liver"
1..4503
/gene="CPSI"
1..4503
/gene="CPSI"
/codon_start=1
/product="carbamoyl phosphate synthetase 1"
/protein_id="CAA75785.1"
/db_xref="GI:3228248"
/db_xref="SWISS-PROT:P31327"
/translation="MTRILTAQVVRVTLTGFTGNTVTAHQKWFSPGRGIRLLSVKAO
TAHIVLEGTAKMGYSFGHSSVAGVYVFNGLGAYPEATDPAKQILITWANPIAG
NGRPDFTALDELGLSKYLESSGKIVSGLLVDYSKDTNHNWLAATKSLQWQLEKVPV
IYGVDTMLTKIIRDKTGLGTEFEGQVDFVFNKONLTAETSKDVKYVKGKNTQ
KYVAVDCGKKNVIRLLVKGAEVHLVFNWHDFTMEYDGLIAGGPGPALAEPLIQ
NVRKLESDRKEPLFGISTGNLTGLAGAKTYKMSMNRQNPVNTNKOATIFD
ONHGYALDNTLPAGMKPLFVNVNDQNEGIMHESKPFPAVFHEPTPGPIDTEYLF
SFESLIKKGKATITTSVLKPALYASRYEYKVLILSGGLSLGQAGEFDYSGSAVK
AKNEENVKVLNPNFASQVINEVGLKQADIVFIPIPFQVTEVINKAQDGLILGM
GGCTALNGVELFKRGVLYKEYGVKLGTSVEMATEDROLFSKLINEIKIAPSPA
VESIEDALKAAADITGYPMIRSAIYALGSLGSGICPNRETLMDSIKAFAMTQILVEK
SVTGKEIEYEVWDADDNCVTYCNMENVDMGVHTGDSVVVAPQOTLSNAEFQMLRR
TSINVVHLGIVGECNIQFALHPTSMKYCIIEVNARLSRSALASKATGYPLAFTAAR
IALGIPLEIKNVYSGTSACPEPSIDYMTKIPRWLDREHGTSSRLSSKMSKSYEV
MALGTFEESQKALRMCHPSIEGFTPLPMNKWPNPNDLRKELSEPSSTRIYAIAR
AIDNNMSDELEKITYIDKWFLYMRDILNMEKTLKGLNSESMTETTLKRAKEIFGSD
KQISCLGLTBAQTRELRLKNTHPWKQIDTLAAETPSVNTLYTVYTGQSHEDYFND
DGMVAVLGGYHIGSVFEDWCASSIRTLQLGKTVVNVNCPNTVSTGTDDECDKL
YFEELSRLIDYIYHQAACGGCIISVGGQIPNNLAVPLKYNKVGKIMGTSPLQIDRAED
RSIESVLDLDELKVAQAPKAVNTLNEALEFAKSDYDPCLLRPSVYLSGSAWYFSED
EMKKFLEBATRVSGEHPVLTKEFEGAREVEMDVGKGRVISHALSEHVEDAGVHSR
DATMLPQTQISQAEIKVKDARAKAFASGPFNPFVQVFLKGVNDVLYIECNLRAS
SPFFVYKTLGVDFIDVATKVMIGENVDEKHLPTDHPITIIDYVAIKAPMFSWPLRD
ADPILRCEMASTGEVACFGEIGIHTAFKLAMLTGTGKIPQKIGILIGIQOQFRPRFLGVA
EOLHNHGFKLFAEATSDMLNANNVPATVPAMPQSGONPSSLIRKLKRDGSDILVI
NLPNNNTFVHDNVIVRTAVDSGIPLLTNQVVKLFAEAVQKSRKVDKSLPHYRQY
SAGKAA"
1631
/gene="CPSI"
/note="mutation in CPSI deficiency"
/replacement="t"
2679
/gene="CPSI"
/note="putative polymorphism (silent)"
/replacement="g"
2838
/gene="CPSI"
/note="ambiguous"
variation
variation
variation

```

```
/db_xref="taxon:9606"
/chromosomes="2"
/map="2q35"
84..196
/gene="CPSI"
/number=36
BASE COUNT      93 a    54 c    48 g    91 t
ORIGIN

Query Match      98.0%; Score 79.4; DB 9; Length 296;
Best Local Similarity 98.8%; Pred. No. 5e-17;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAACGCCCAACAATGTCCTGCCAACCCAGTGGCATGGCCGTCT 60
   |||||||
Db 99 GCCACATCAGACTGGCTCAACGCCCAACAATGTCCTGCCAACCCAGTGGCATGGCCGTCT 158
   |||||||

QY 61 CAAGAAGGACAGAATCCCGCAG 81
   |||||||
Db 159 CAAGAAGGACAGAATCCCGCAG 179
   |||||||

RESULT 8
LOCUS      AR184418              495 bp    DNA    linear    PAT 20-APR-2002
DEFINITION Sequence 5 from patent US 6346382.
ACCESSION  AR184418
VERSION     AR184418.1 GI:20230383
KEYWORDS   .
SOURCE      Unknown.
ORGANISM    Unknown.
REFERENCE   1 (bases 1 to 495)
AUTHORS     Summar,M.L. and Christman,B.W.
TITLE       Human carbonyl phosphate synthetase I polymorphism and diagnostic
            methods related thereto
JOURNAL     Patent: US 6346382-A 5 12-FEB-2002;
            Location/Qualifiers
            1..495
            /organism="unknown"
BASE COUNT  158 a    96 c    85 g    155 t    1 others
ORIGIN

Query Match      98.0%; Score 79.4; DB 6; Length 495;
Best Local Similarity 98.8%; Pred. No. 4.7e-17;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAACGCCCAACAATGTCCTGCCAACCCAGTGGCATGGCCGTCT 60
   |||||||
Db 126 GCCACATCAGACTGGCTCAACGCCCAACAATGTCCTGCCAACCCAGTGGCATGGCCGTCT 185
   |||||||

QY 61 CAAGAAGGACAGAATCCCGCAG 81
   |||||||
Db 186 CAAGAAGGACAGAATCCCGCAG 206
   |||||||

RESULT 9
LOCUS      AR272381              531 bp    DNA    linear    PAT 10-APR-2003
DEFINITION Sequence 93 from patent US 6504010.
ACCESSION  AR272381
VERSION     AR272381.1 GI:29704266
KEYWORDS   .
SOURCE      Unknown.
ORGANISM    Unknown.
REFERENCE   1 (bases 1 to 531)
AUTHORS     Wang,T., Bangur,C.S., Lodes,M.J., Fanger,G.R., Vedvick,T.S.,
            Carter,D., Retter,M.W., Mannion,J., and Fan,L.
TITLE       Compositions and methods for the therapy and diagnosis of lung
            cancer
JOURNAL     Patent: US 6504010-A 93 07-JAN-2003;
            Location/Qualifiers
            1..531
            /organism="unknown"
BASE COUNT  157 a    139 c    108 g    127 t
ORIGIN

Query Match      98.0%; Score 79.4; DB 6; Length 531;
Best Local Similarity 98.8%; Pred. No. 4.7e-17;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAACGCCCAACAATGTCCTGCCAACCCAGTGGCATGGCCGTCT 60
   |||||||
Db 157 GCCACATCAGACTGGCTCAACGCCCAACAATGTCCTGCCAACCCAGTGGCATGGCCGTCT 216
   |||||||

QY 61 CAAGAAGGACAGAATCCCGCAG 81
   |||||||
Db 217 CAAGAAGGACAGAATCCCGCAG 237
   |||||||

RESULT 11
LOCUS      AX062466              531 bp    DNA    linear    PAT 24-JAN-2001
DEFINITION Sequence 93 from Patent WO0100828.
ACCESSION  AX062466
VERSION     AX062466.1 GI:12540341
KEYWORDS   .
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
REFERENCE   1
AUTHORS     Wang,T., Bangur,C.S., Lodes,M.J., Fanger,G.R., Vedvick,T.S.,
            Carter,D., Retter,M.W. and Mannion,J.
TITLE       Compositions and methods for the therapy and diagnosis of lung
            cancer
JOURNAL     Patent: WO 0100828-A 93 04-JAN-2001;
            CORIXA CORPORATION (US)
            Location/Qualifiers
            1..531
            /organism="unknown"
BASE COUNT  157 a    139 c    108 g    127 t
ORIGIN

Query Match      98.0%; Score 79.4; DB 6; Length 531;
Best Local Similarity 98.8%; Pred. No. 4.7e-17;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAACGCCCAACAATGTCCTGCCAACCCAGTGGCATGGCCGTCT 60
   |||||||
Db 157 GCCACATCAGACTGGCTCAACGCCCAACAATGTCCTGCCAACCCAGTGGCATGGCCGTCT 216
   |||||||

QY 61 CAAGAAGGACAGAATCCCGCAG 81
   |||||||
Db 217 CAAGAAGGACAGAATCCCGCAG 237
   |||||||
```

```
source      1..531
            /organism="unknown"
BASE COUNT  157 a    139 c    108 g    127 t
ORIGIN

Query Match      98.0%; Score 79.4; DB 6; Length 531;
Best Local Similarity 98.8%; Pred. No. 4.7e-17;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAACGCCCAACAATGTCCTGCCAACCCAGTGGCATGGCCGTCT 60
   |||||||
Db 157 GCCACATCAGACTGGCTCAACGCCCAACAATGTCCTGCCAACCCAGTGGCATGGCCGTCT 216
   |||||||

QY 61 CAAGAAGGACAGAATCCCGCAG 81
   |||||||
Db 217 CAAGAAGGACAGAATCCCGCAG 237
   |||||||

RESULT 10
LOCUS      AR275962              531 bp    DNA    linear    PAT 10-APR-2003
DEFINITION Sequence 93 from patent US 6509448.
ACCESSION  AR275962
VERSION     AR275962.1 GI:29709607
KEYWORDS   .
SOURCE      Unknown.
ORGANISM    Unknown.
REFERENCE   1 (bases 1 to 531)
AUTHORS     Wang,T., Bangur,C.S., Lodes,M.J., Fanger,G.R., Vedvick,T.S.,
            Carter,D., Retter,M.W., Mannion,J., Fan,L. and Wang,A.
TITLE       Compositions and methods for the therapy and diagnosis of lung
            cancer
JOURNAL     Patent: US 6509448-A 93 21-JAN-2003;
            Location/Qualifiers
            1..531
            /organism="unknown"
BASE COUNT  157 a    139 c    108 g    127 t
ORIGIN

Query Match      98.0%; Score 79.4; DB 6; Length 531;
Best Local Similarity 98.8%; Pred. No. 4.7e-17;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACATCAGACTGGCTCAACGCCCAACAATGTCCTGCCAACCCAGTGGCATGGCCGTCT 60
   |||||||
Db 157 GCCACATCAGACTGGCTCAACGCCCAACAATGTCCTGCCAACCCAGTGGCATGGCCGTCT 216
   |||||||

QY 61 CAAGAAGGACAGAATCCCGCAG 81
   |||||||
Db 217 CAAGAAGGACAGAATCCCGCAG 237
   |||||||

RESULT 11
LOCUS      AX062466              531 bp    DNA    linear    PAT 24-JAN-2001
DEFINITION Sequence 93 from Patent WO0100828.
ACCESSION  AX062466
VERSION     AX062466.1 GI:12540341
KEYWORDS   .
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
REFERENCE   1
AUTHORS     Wang,T., Bangur,C.S., Lodes,M.J., Fanger,G.R., Vedvick,T.S.,
            Carter,D., Retter,M.W. and Mannion,J.
TITLE       Compositions and methods for the therapy and diagnosis of lung
            cancer
JOURNAL     Patent: WO 0100828-A 93 04-JAN-2001;
            CORIXA CORPORATION (US)
            Location/Qualifiers
            1..531
            /organism="unknown"
BASE COUNT  157 a    139 c    108 g    127 t
ORIGIN
```

```

NLPNNNTKVFHDNTVIRRTAVDSGIGLLTNFOVTKLPFAEAVQSRKVDKSLFHYRQY
SAGKAA"
BASE COUNT      1484 a      1102 c      1207 g      1422 t
ORIGIN

Query Match      100.0%; Score 81; DB 9; Length 5215;
Best Local Similarity 100.0%; Pred. No. 1e-17;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GCCACATCAGACTGGCTCAAGCCCAACAATGTCCTGCCAACCCAGTGGCATGGCCGTCT 60
        |||||||
Db      4295 GCCACATCAGACTGGCTCAAGCCCAACAATGTCCTGCCAACCCAGTGGCATGGCCGTCT 4354

QY      61 CAAGAAGGACAGAATCCCGAGC 81
        |||||||
Db      4355 CAAGAAGGACAGAATCCCGAGC 4375

RESULT 4
LOCUS      AR184416                      5761 bp      DNA      linear      PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6346382.
ACCESSION  AR184416
VERSION     AR184416.1 GI:20230381
KEYWORDS   Unknown.
SOURCE     Unknown.
ORGANISM   Unclassified.
REFERENCE  1 (bases 1 to 5761)
AUTHORS    Summar,M.L. and Christman,B.W.
TITLE      Human carbamyl phosphate synthetase I polymorphism and diagnostic
           methods related thereto
JOURNAL    Patent: US 6346382-A 1 12-FEB-2002;
FEATURES   Location/Qualifiers
           source
             1..5761
             /organism="unknown"
BASE COUNT  1625 a      1209 c      1309 g      1618 t
ORIGIN

Query Match      100.0%; Score 81; DB 6; Length 5761;
Best Local Similarity 100.0%; Pred. No. 1e-17;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GCCACATCAGACTGGCTCAAGCCCAACAATGTCCTGCCAACCCAGTGGCATGGCCGTCT 60
        |||||||
Db      4300 GCCACATCAGACTGGCTCAAGCCCAACAATGTCCTGCCAACCCAGTGGCATGGCCGTCT 4359

QY      61 CAAGAAGGACAGAATCCCGAGC 81
        |||||||
Db      4360 CAAGAAGGACAGAATCCCGAGC 4380

RESULT 5
LOCUS      AR184424                      5761 bp      DNA      linear      PAT 20-APR-2002
DEFINITION Sequence 11 from patent US 6346382.
ACCESSION  AR184424
VERSION     AR184424.1 GI:20230389
KEYWORDS   Unknown.
SOURCE     Unknown.
ORGANISM   Unclassified.
REFERENCE  1 (bases 1 to 5761)
AUTHORS    Summar,M.L. and Christman,B.W.
TITLE      Human carbamyl phosphate synthetase I polymorphism and diagnostic
           methods related thereto
JOURNAL    Patent: US 6346382-A 11 12-FEB-2002;
FEATURES   Location/Qualifiers
           source
             1..5761
             /organism="unknown"
BASE COUNT  1625 a      1209 c      1309 g      1618 t
ORIGIN

NLPNNNTKVFHDNTVIRRTAVDSGIGLLTNFOVTKLPFAEAVQSRKVDKSLFHYRQY
SAGKAA"
BASE COUNT      1484 a      1102 c      1207 g      1422 t
ORIGIN

Query Match      100.0%; Score 81; DB 9; Length 5215;
Best Local Similarity 100.0%; Pred. No. 1e-17;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GCCACATCAGACTGGCTCAAGCCCAACAATGTCCTGCCAACCCAGTGGCATGGCCGTCT 60
        |||||||
Db      4295 GCCACATCAGACTGGCTCAAGCCCAACAATGTCCTGCCAACCCAGTGGCATGGCCGTCT 4354

QY      61 CAAGAAGGACAGAATCCCGAGC 81
        |||||||
Db      4355 CAAGAAGGACAGAATCCCGAGC 4375

RESULT 4
LOCUS      AR184416                      5761 bp      DNA      linear      PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6346382.
ACCESSION  AR184416
VERSION     AR184416.1 GI:20230381
KEYWORDS   Unknown.
SOURCE     Unknown.
ORGANISM   Unclassified.
REFERENCE  1 (bases 1 to 5761)
AUTHORS    Summar,M.L. and Christman,B.W.
TITLE      Human carbamyl phosphate synthetase I polymorphism and diagnostic
           methods related thereto
JOURNAL    Patent: US 6346382-A 1 12-FEB-2002;
FEATURES   Location/Qualifiers
           source
             1..5761
             /organism="unknown"
BASE COUNT  1625 a      1209 c      1309 g      1618 t
ORIGIN

Query Match      100.0%; Score 81; DB 6; Length 5761;
Best Local Similarity 100.0%; Pred. No. 1e-17;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GCCACATCAGACTGGCTCAAGCCCAACAATGTCCTGCCAACCCAGTGGCATGGCCGTCT 60
        |||||||
Db      4300 GCCACATCAGACTGGCTCAAGCCCAACAATGTCCTGCCAACCCAGTGGCATGGCCGTCT 4359

QY      61 CAAGAAGGACAGAATCCCGAGC 81
        |||||||
Db      4360 CAAGAAGGACAGAATCCCGAGC 4380

RESULT 5
LOCUS      AR184424                      5761 bp      DNA      linear      PAT 20-APR-2002
DEFINITION Sequence 11 from patent US 6346382.
ACCESSION  AR184424
VERSION     AR184424.1 GI:20230389
KEYWORDS   Unknown.
SOURCE     Unknown.
ORGANISM   Unclassified.
REFERENCE  1 (bases 1 to 5761)
AUTHORS    Summar,M.L. and Christman,B.W.
TITLE      Human carbamyl phosphate synthetase I polymorphism and diagnostic
           methods related thereto
JOURNAL    Patent: US 6346382-A 11 12-FEB-2002;
FEATURES   Location/Qualifiers
           source
             1..5761
             /organism="unknown"
BASE COUNT  1625 a      1209 c      1309 g      1618 t
ORIGIN
```

```

Query Match      100.0%; Score 81; DB 6; Length 5761;
Best Local Similarity 100.0%; Pred. No. 1e-17;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GCCACATCAGACTGGCTCAAGCCCAACAATGTCCTGCCAACCCAGTGGCATGGCCGTCT 60
        |||||||
Db      4300 GCCACATCAGACTGGCTCAAGCCCAACAATGTCCTGCCAACCCAGTGGCATGGCCGTCT 4359

QY      61 CAAGAAGGACAGAATCCCGAGC 81
        |||||||
Db      4360 CAAGAAGGACAGAATCCCGAGC 4380

RESULT 6
LOCUS      AR184425                      5762 bp      DNA      linear      PAT 20-APR-2002
DEFINITION Sequence 13 from patent US 6346382.
ACCESSION  AR184425
VERSION     AR184425.1 GI:20230390
KEYWORDS   Unknown.
SOURCE     Unknown.
ORGANISM   Unclassified.
REFERENCE  1 (bases 1 to 5762)
AUTHORS    Summar,M.L. and Christman,B.W.
TITLE      Human carbamyl phosphate synthetase I polymorphism and diagnostic
           methods related thereto
JOURNAL    Patent: US 6346382-A 13 12-FEB-2002;
FEATURES   Location/Qualifiers
           source
             1..5762
             /organism="unknown"
BASE COUNT  1626 a      1210 c      1308 g      1618 t
ORIGIN

Query Match      100.0%; Score 81; DB 6; Length 5762;
Best Local Similarity 100.0%; Pred. No. 1e-17;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GCCACATCAGACTGGCTCAAGCCCAACAATGTCCTGCCAACCCAGTGGCATGGCCGTCT 60
        |||||||
Db      4300 GCCACATCAGACTGGCTCAAGCCCAACAATGTCCTGCCAACCCAGTGGCATGGCCGTCT 4359

QY      61 CAAGAAGGACAGAATCCCGAGC 81
        |||||||
Db      4360 CAAGAAGGACAGAATCCCGAGC 4380

RESULT 7
LOCUS      AY166970S36                  296 bp      DNA      linear      PRI 28-FEB-2003
DEFINITION Homo sapiens carbanoylphosphate synthetase (CPSI) gene, exon 36;
           nuclear gene for mitochondrial product.
ACCESSION  AY167005
VERSION     AY167005.1 GI:28625483
KEYWORDS   36 of 38
SEGMENT    Homo sapiens (human)
SOURCE     Homo sapiens
           Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 296)
AUTHORS    Fughini,S., Morrone,A. and Zammarchi,E.
TITLE      Human carbanoylphosphate synthetase (CPS I) gene organization
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 296)
AUTHORS    Fughini,S., Morrone,A. and Zammarchi,E.
TITLE      Direct Submission
JOURNAL    Submitted (16-OCT-2002) Dep. of Paediatrics, University of
           Florence, Via Luca Giordano 13, Florence 50132, Italy
FEATURES   Location/Qualifiers
           source
             1..296
             /organism="Homo sapiens"
             /mol_type="genomic DNA"
```

```

FEATURES
  source
    GENE LOGIC INC (US)
      Location/Qualifiers
        1..5215
          /organism="Homo sapiens"
          /mol_type="genomic DNA"
          /db_xref="taxon:9606"
          /note="EMBL/GenBank Accession No. D90282"
BASE COUNT    1484 a 1102 c 1207 g 1422 t
ORIGIN
  Query Match      100.0%; Score 81; DB 6; Length 5215;
  Best Local Similarity 100.0%; Pred. No. 1e-17;
  Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTGCCAACCCAGTGGCATTGCCCGTCT 60
    |||||||
Db 4295 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTGCCAACCCAGTGGCATTGCCCGTCT 4354
QY 61 CAAGAAGGACAGATCCCGAGC 81
    |||||||
Db 4355 CAAGAAGGACAGATCCCGAGC 4375

RESULT 2
LOCUS      E04065      5215 bp      RNA      linear      PAT 29-SEP-1997
DEFINITION CDNA encoding human carbamyl phosphate synthetase 1.
ACCESSION  E04065
VERSION    E04065.1 GI:2172275
KEYWORDS  JP 1992335889-A/1.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 5215)
AUTHORS   Haraguchi, Y., Mori, M. and Matsuda, I.
TITLE     GENE AND DNA PROVE TO USED THEREFOR
JOURNAL   Patent: JP 1992335889-A 1 24-NOV-1992;
COMMENT   HARAGUCHI YOGO, MORI MASATAKA, MATSUDA ICHIRO
          OS Homo sapiens (human)
          PN JP 1992335889-A/1
          PD 24-NOV-1992
          PF 09-MAY-1991 JP 1991135902
          PI HARAGUCHI YOGO, MORI MASATAKA, MATSUDA ICHIRO PC
          CL2N15/52,C07H21/04,CL2N15/11,CL2Q1/68//A61B10/00; CC
          strandedness: Double;
          CC topology: Linear;
          CC hypothetical: No;
          CC anti-sense: No;
          CC *source: tissue_type=Liver;
          FH Key      Location/Qualifiers
          FH          1..118
          FT 5'UTR
          FT CDS
          FT 3'UTR
          Location/Qualifiers
            1..5215
              /organism="Homo sapiens"
              /mol_type="genomic RNA"
              /db_xref="taxon:9606"
BASE COUNT    1484 a 1102 c 1207 g 1422 t
ORIGIN
  Query Match      100.0%; Score 81; DB 6; Length 5215;
  Best Local Similarity 100.0%; Pred. No. 1e-17;
  Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTGCCAACCCAGTGGCATTGCCCGTCT 60
    |||||||
Db 4295 GCCACATCAGACTGGCTCAACGCCAACAAATGTCCTGCCAACCCAGTGGCATTGCCCGTCT 4354

```

```

QY 61 CAAGAAGGACAGATCCCGAGC 81
    |||||||
Db 4355 CAAGAAGGACAGATCCCGAGC 4375

RESULT 3
LOCUS      HUMCPSI      5215 bp      mRNA      linear      PRI 29-MAY-2002
DEFINITION Homo sapiens mRNA for carbamyl phosphate synthetase 1, complete
            cds.
ACCESSION  D90282
VERSION    D90282.1 GI:219552
KEYWORDS
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 5215)
AUTHORS   Haraguchi, Y., Uchino, T., Takiguchi, M., Endo, F., Mori, M. and
            Matsuda, I.
TITLE     Cloning and sequence of a cDNA encoding human carbamyl phosphate
            synthetase 1: molecular analysis of hyperammonemia
JOURNAL   Gene 107 (2), 335-340 (1991)
MEDLINE   92084128
PUBMED    1840546
COMMENT   Submitted (18-JAN-1991) to DDBJ by:
            Yougo Haraguchi
            Department of Pediatrics
            Kumamoto University Medical School
            1-1-1 Honjo
            Kumamoto 860
            Japan
            Phone: 096-344-2111 x5654
            Fax: 096-366-3471.
            Location/Qualifiers
              1..5215
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /chromosome="2"
                /tissue_type="liver"
                /note="short arm"
              119..4621
                /EC_number="6.3.4.16"
                /codon_start=1
                /product="carbamyl phosphate synthetase I"
                /protein_id="BAA14328.1"
                /db_xref="GI:219553"
                /translation="MTRILTAFKVVRVRLKTFGFTNVTAKQKWFSPRGIRLLSVKQAQ
                TAHVLEDTGKMGYSFGHPSVAGVEVNTGLGYPALITDPAVKGOILTMANPLIG
                NGGAPDTSLDELGLSKYLENGIKVSGLLVLDYSDYNHNLATKSLGQWLQEKVPA
                IYGVDFMLTKIIRDKGTMLKIEFGQPVDFDPNKNQIAETVSKDKVYIGKGNPT
                KVAVDGCIKNNVIRLLYKRGAEHLVFWNHDFTKMEYDGLIAGGPNPALAEPLIQ
                NVQKILESDEKPEFLFGISTGNLITLGAAGAKTYKSMANRQNGQVFLNITNKQAFITA
                QNSHLKADNTLPAQKPLFVNVNDQTNEGIMHESKPFVAFVPHPEVTPGPTDEYLFDF
                SFESLKLKATTTISVLKPAALVASRVEYSKVLILGSGGLSIGAGGEEDYSGSQAVK
                AMKEENVKTVLNPNIASVQINEVGLQAQDTVYFLPITPQVYTVIKAEQDGLILGM
                GGQFALMCGVELFKRGVLKEIGVKVLGTSVESINATEDQLFSDKLNEINEKIAPSEFA
                VESIEDALKAADTIQIPVHIRISAYALGGLGSGICPNRETLMDSLTKAFATNQLLVEK
                SVTGWKEIEYEVVRDADDNCTVCMENVDAMGVHTGDSVVVAPAAQTLSNAEFQMLER
                TSNVNRHLGIYGECEIOFALHPTSMGYCIIEVNAKSPNSALAKTGYPLAFIAAK
                IALGILPLGIKNVSVSKTSACFEPSLDYMTKIPMDLDRPHGTSRIGSGSMKSVGEV
                MAIGRPFESFQKALRMCHPSIEGTFPLPMKKEWPSNLDLKLSESPSTRYIAAK
                AIDNMSLEIEKLTIIDKFLYINRDLINKEATLKLNSMTEETLKRKEIGFSD
                KQISCLGTEAQTRELKRNHPVKQIDTLAAEYFSVNYLYVTYINGQHDVNFDF
                DHGMVILGGCPYHIGSVFDFQCAVSSIRTQLGKTKVYVNCNPETVSTDFDECDKL
                YFESLSRLDLIDYHOEACGGCIIISVGQIPNNIAVPLKNGVKIMGTSPLOIDRADF
                RSIFSALVELDKVQAPKWKAYNTLNEALEFAKSDYDPCLLRPSYVLSGSANVVSSED
                EMKKFDEEATRVSQATPVVLTFRFVEGAREVDVAGKGVISHAISHEVDAGVHSE
                NATMLPTQTIQSQAEIKVDAKRIAKAFISGPNVQFLYKNGDVLVINECHLRASR
                SFPSSKTLGVDFIDVATKVLIGENVDEKHLPLTDHPITIPDYVAIKAPFNSWPLRDR
                ADPLRCEMASTGEVACFEGEHTAFKMLSTGFKIPKQILITIGIQSFRPRFLGVA
                EQLHNEGFKLFATEATSDNLNANNVPANVPANFSPQEGQNPSSLSIRKLIRDGSDILVI

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: September 12, 2003, 22:25:46 ; Search time 1416.5 seconds
(without alignments)
2339.343 Million cell updates/sec

Title: US-09-585-077c-1_copy_4300_4380
Perfect score: 81
Sequence: 1 gccacatcagactggctcaa.....aagaaggacagaatccaccg 81

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 2888711 seqs, 2045481386 residues
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.in.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.man.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	81	100.0	5215	6	AX409122	Sequence
2	81	100.0	5215	6	E04065	E04065 cDNA encodi
3	81	100.0	5215	9	H0MCP51	D90282 Homo sapien
4	81	100.0	5761	6	AR184416	AR184416 Sequence
5	81	100.0	5761	6	AR184424	AR184424 Sequence
6	81	100.0	5762	6	AR184425	AR184425 Sequence
7	79.4	98.0	296	9	AY166970S36	AY167005 Homo sapi
8	79.4	98.0	495	6	AR184418	AR184418 Sequence
9	79.4	98.0	531	6	AR272381	AR272381 Sequence
10	79.4	98.0	531	6	AR275962	AR275962 Sequence
11	79.4	98.0	531	6	AX062466	AX062466 Sequence
12	79.4	98.0	531	6	AX367383	AX367383 Sequence
13	79.4	98.0	4503	9	HSCPSIMUT	Y15793 Homo sapien
14	79.4	98.0	5761	6	AR184417	AR184417 Sequence
15	79.4	98.0	5761	9	AF154830	AF154830 Homo sapi
16	79.4	98.0	121959	9	AF536523	AF536523 Homo sapi
17	79.4	98.0	183696	9	AC008172	AC008172 Homo sapi
18	73	90.1	133	10	RNCP5IX36	Z27487 R.norvegicu
19	73	90.1	147	10	RNCP5I11	M12327 Rattus norv
20	73	90.1	229369	2	AC106119	AC106119 Rattus no
21	69.8	86.2	255581	2	AC101854	AC101854 Mus muscu
22	69	85.2	816	6	BD091111	BD091111 DNA chip
23	47	58.0	4716	5	RCU05193	U05193 Rana catesb
24	46	56.8	580	6	AX525511	AX525511 Sequence
25	41	50.6	4597	5	OMU65893	U65893 Oncorhynch
26	38.4	47.4	4717	5	AF006491	AF006491 Micropter
27	33	40.7	63058	2	AC133885	AC133885 Homo sapi
28	30.6	37.8	31	6	AX249689	AX249689 Sequence
29	30	37.0	6001	5	SQUCARPSYN	L31362 Squalus aca
30	29.6	36.5	4755	5	AF119250	AF119250 Oreochrom
31	29.6	36.5	222983	2	AC129695	AC129695 Rattus no
32	29.6	36.5	253973	2	AC094544	AC094544 Rattus no
33	29.6	36.5	283158	2	AC137193	AC137193 Rattus no
34	29.6	36.5	320205	2	AC106178	AC106178 Rattus no
35	29.4	36.3	36307	9	HS366D1	Z97986 Human DNA s
36	29.4	36.3	256073	9	AE006454	AE006454 Homo sapi
37	29.2	36.0	154736	9	AL139246	AL139246 Human DNA
38	29	35.8	1101	11	G67204	G67204 Xq4133 KWOK
39	29	35.8	141672	9	HS142F18	AL031073 Human DNA
40	28	34.6	63058	2	AC133885	AC133885 Homo sapi
41	28	34.6	72620	2	AC121127	AC121127 Mus muscu
42	28	34.6	161065	9	AC124799	AC124799 Homo sapi
43	28	34.6	161190	9	AC005137	AC005137 Human Chr
44	28	34.6	234487	2	AC094748	AC094748 Rattus no
45	27.8	34.3	5989	5	AF169248	AF169248 Opsanus b

ALIGNMENTS

RESULT 1
AX409122
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL

AX409122
Sequence 1769 from Patent WO0229103.
AX409122
AX409122.1 GI:21441827
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.
Gene expression profiles in liver cancer
Patent: WO 0229103-A 1769 11-APR-2002;

AX409122
Sequence 1769 from Patent WO0229103.
AX409122
AX409122.1 GI:21441827
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.
Gene expression profiles in liver cancer
Patent: WO 0229103-A 1769 11-APR-2002;